Australian Education Review

Reforming Educational Assessment:

Imperatives, principles and challenges

Geoff N Masters

Australian Council *for* Educational Research

ACER

Australian Education Review

Reforming Educational Assessment: Imperatives, principles and challenges

Geoff N Masters

Australian Council *for* Educational Research

First published 2013
by ACER Press
Australian Council *for* Educational Research
19 Prospect Hill Road, Camberwell, Victoria, 3124

Series Editor: Suzanne Mellor
Copyedited by Carolyn Glascodine
Typeset by ACER Project Publishing
Printed by BPA Print Group

National Library of Australia Cataloguing-in-Publication entry

Author:	Masters, Geofferey N., 1952-
Title:	Reforming education assessment : imperatives, principles and challenges / Geoff N Masters.
ISBN:	9781742862026 (pbk.)
Series:	Australian education review ; no. 57.
Notes:	Includes bibliographical references.
Subjects:	Educational tests and measurements--Australia. Students--Rating of--Australia. Education--Evaluation--Australia. Education--Aims and objectives--Australia. Teaching--Evaluation--Australia.
Other Authors/Contributors:	Australian Council for Educational Research.
Dewey Number:	371.26013
Visit our website:	www.acer.edu.au/aer

Acknowledgements for cover image:
Michael Anderson, Paramount Studios

Foreword

Today what happens in education is of general public interest and receives much media attention. As a result, many education goals are short-term responses to a range of political and public aspirations, rather than being developed in the light of careful consideration of what is deliverable as educational objectives. Global education imperatives make it necessary for each country to focus attention on how best to achieve and measure greater education inclusion, higher-quality learning and better attainment rates in order to meet social and economic needs (Freedman, 2010).

Around the world government reforms to education systems have been driven by national and international assessment data (Kellaghan & Greaney, 2001). Consequently, whole education systems are being judged on the outcomes of student performance on standardised tests and public examinations. International test results, being understood and used as performance indicators of the health of education systems and institutions, is now a well established aspect of the public and academic discourse about education. Given the limited scope of their coverage of delivered curriculum, such influence is disproportionate to any intrinsic value they may have as a proxy for educational outcomes.

Test results which are lower than expected relative to other countries, or lower than on previous occasions of testing, are taken as an indication of system failure. This creates pressure at every level of education systems. Impacts on education policy and practice range from country level, where judgements about skill level and relative international strength of human capital development are made, down to the school and classroom level, where inferences are made about school and teacher effectiveness (Stanley, 2012).

Recently the governments of the United Kingdom, the United States of America and Australia have been prompted to introduce reform and improvement policies in response to change in their country's relative performance on PISA. Their explicit aim is to raise their ranking relative to other countries. Setting national or local education targets based on rank position on international tests, rather than on specific standards achieved, flies in the face of modern assessment practice, which is standards-based. Rank position is inappropriate as a goal for improvement, as national rankings can be influenced by quite small differences in student cohort scores. Such differences may not, in themselves, represent meaningful learning and/or skill differences in standards attained, which should be the main focus. Despite these concerns, rank position appeals to those who see improving educational outcomes as a competitive sport between nations.

Student performance on current international tests is not broadly enough based to assume the sort of bottom-line significance for education that, for example, GDP growth assumes for comparing the health of economies. However, more comprehensive assessment of education outcomes may well be too expensive to obtain. Regrettably it is likely that the education

community will need to accept the current tests as the ones that will continue to be used as indicators of educational growth/health. Given this, there is a clear need for educators and policymakers to have a more informed understanding of the strengths and weaknesses involved in public policy approaches which interpret current assessment data as a performance indicator.

While it is important to acknowledge that assessment has commonly been given many roles in contemporary education practice (Newton, 2007), fundamentally assessment is based on student responses and its primary purpose should be of benefit to students. Geoff Masters in this review makes an excellent argument that both assessment itself, and our thinking about assessment, are in need of reform.

How people learn, drawing on insights from developmental psychology, cognitive science and neuroscience is at the heart of the emerging focus on evidence-based practice in education. There is an urgent need for all educators, as well as stakeholders, to consider how practice in assessment can be improved by this knowledge and how assessment data should be interpreted. As a starter, there is a need to be clear about what we mean by assessment in education.

The current educational literature is full of different positions and terminology, which teachers and students can find confusing. In this timely review Masters re-centres the discussion about assessment on current understanding of how students learn, and how their knowledge and understanding grows. It is helpful to focus on his simple unifying principle: *The fundamental purpose of assessment is to establish where learners are in their learning at the time of assessment.*

The power of this principle is that it directly links assessment directly to student learning and to the consideration of evidence about growth and development. It is in line with a broader move towards recognising that education is about personal learning and that formal education needs to embrace a better understanding of other learning occurring outside the formal classroom setting.

For all those with a direct interest in the teaching/learning process, assessment needs to be about evidence of progress in the growth of knowledge, understanding and skills. Unfortunately, many debates among educators focus on mode of assessment rather than on consistency or usefulness of evidence.

Evidence can take many forms, but regardless of the form, it needs to be considered in terms of how well it satisfies needs for practicality, fairness, validity and whether it provides feedback to assist the next step in the developmental pathway for an individual. Timely feedback is essential to assist learning (Hattie & Timperley, 2007).

Different sources of evidence about student growth should converge. For example, if in a particular case there are different attainment signals coming from external tests, compared to classroom observation, rather than writing off one source there is value in adopting a forensic approach to understanding why such a discrepancy has occurred. The end product of such analysis should lead to more effective understanding of student learning.

By defining educational assessment with reference to growth, there are certain challenges to current practices. If curriculum requirements are not organised with respect to developmental outcomes that clarify expected learning pathways, then teaching programs are unlikely to yield evidence of depth of learning. The consequences of such an approach to assessment is a renegotiating of the processes of curriculum, teaching and assessment towards a holistic emphasis on how growth occurs and on what evidence should be gathered to show that it is occurring.

As Krajcik (2011) has pointed out 'most curriculum materials that currently exist focus primarily on impoverished ideas about student learning or are based on no model of learning at all, and few, if any, follow a development perspective' (p. 156). To embody a growth perspective in assessment requires disciplined thinking and careful research, in order to ensure that more than surface learning is being encouraged by the teaching/learning program at school level. There needs to be a shift away from judging success in just learning specific content, towards judging how students are developing in understanding.

It takes time and resources to develop research-based learning progressions and so far most attention to such development has been in areas such as literacy, numeracy and science (e.g. Black, Wilson & Yao, 2011; McNamara & Hill, 2011). These curriculum areas have been given special attention because of their core nature and apparent tractability to a developmental pathway. Even so, consensus has taken time, and without the stimulus of the IEA and OECD testing programs, even less progress may have been made.

At the heart of the Masters proposal for assessment reform is the set of principles underpinning his model of a Learning Assessment System. The key consequence of its implementation would be a shift away from an age-based, lock-step process of classroom activity, towards one focused on the developmental needs of each student derived from evidence-gathering and observation with respect to empirically based learning progressions.

To assist the teacher in achieving the requirement for more diagnostic understanding of each student in their learning progress, digital technologies have the potential to provide varied assessment tasks with useful feedback customised to individual developmental levels. One senses that there is a tsunami of change underway from educational use of such technologies. They can be harnessed in the central reform movement away from 'assessment as judging' to 'assessment as understanding'.

The challenges inherent in the reforms advocated in this scholarly and balanced review and critique of current assessment may seem overwhelming to some stakeholders. There will need to be significant changes to current practice. Curriculum and assessment practitioners will need to work more closely together and gain better understandings of current learning and developmental research and its implications for each curriculum domain.

Often the organisational structures of education system authorities do not facilitate such cooperation and learning between specialists. There will need to be constructive inputs to curriculum change. Curriculum designers should not see assessment as being 'the tail that wags the curriculum dog' or expect that psychometricians can sort out problems with poorly constructed test items. Considerable investment will be needed for the development of learning progression frameworks across different domains, for teacher professional development and public awareness. Successful systems invest wisely in improvement.

To achieve these reforms will require resolve and political will. Australia has a proud tradition of educational success and has a well-deserved international recognition. If we want it to continue, then the Masters proposals for reform must be taken seriously. His review should be required reading for all involved with our education system. Given that the reform proposals are founded on a large body of evidence, and resonate with individual needs for a purposeful and rewarding learning experience, we should be sanguine about them occurring.

Gordon Stanley is Honorary Professor of Education at the University of Sydney and Emeritus Professor of Psychology at the University of Melbourne. Formerly President of the NSW Board of Studies, he was the inaugural Pearson Professor of Educational Assessment at the University of Oxford, where he is now Honorary Senior Research Fellow.

References

Black, P., Wilson, M., & Yao, S. (2011): Road maps for learning: A guide to the navigation of learning progressions, measurement. *Interdisciplinary Research and Perspectives*, *9* (2–3), 71–112.

Freedman, G. 2010. *Unlocking the global education imperative*. Accessed February 3, 2013, from http://www.blackboard.com/CMSPages/GetFile.aspx?guid=6032f8df-b6ba-4510-81d2-3198459529dc

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Kellaghan, T., & Greaney, V. (2001). *Using assessment to improve the quality of education*. Paris: UNESCO.

Krajcik, J. (2011). Learning progressions provide road maps for the development and validity of assessments and curriculum materials, measurement. *Interdisciplinary Research and Perspectives*, *9* (2–3), 155–158.

McNamara, T., & Hill, K. (2012): A response from languages, measurement. *Interdisciplinary Research and Perspectives*, *10*(3), 176–183.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, *14*(2), 149–170.

Stanley, G. (2012). *The changing role of educational assessment: From global to local impacts*. Report of an Independent Schools Queensland Think Tank Seminar. Accessed February 3, 2013, from http://www.aisq.qld.edu.au/files/files/whatsnew/Changing%20Role%20of%20Educational%20Assessment.pdf

Contents

section 1

Introduction to the review

This review addresses the role of assessment in education. It observes that the field of educational assessment is currently divided and in disarray. Fault lines fragment the field into differing, and often competing philosophies, methods and approaches. At the same time, there are unprecedented external pressures for assessment reform. These pressures include the following: the need for better information to guide and evaluate educational decision-making; advances being made in understandings of human learning; calls for greater emphasis on the development of a broader range of life skills and attributes; and changes in where and how learning takes place, particularly resulting from advances in technology.

This review had its genesis in an ACER research conference in Perth in 2009 on the topic of assessment and student learning. Strong themes emerging from papers presented at the conference include: the growing need for quality assessment information to inform decision-making at all levels of education; the desirability of assessments that explore and provide insights into where students are in their learning; and the importance of ensuring that assessment and reporting processes are consistent with what is known about the conditions that support successful learning. This review builds on, and attempts to extend, that seminal discussion.

Assessment reform

Educational assessment, as a field, has been subject to increasing pressures to reform over recent decades. These pressures have come from quite different directions. The first of these pressures has been the demand for better data to inform educational decision-making, particularly by governments and education systems, but also by school leaders, parents and classroom teachers. Second, advances in our understanding of human learning have highlighted inconsistencies between many traditional assessment and reporting practices and what is now known about the general conditions that promote successful learning. Third, there has been growing recognition within the business and education communities of the need to develop assessment methods for a broader range of skills and attributes necessary for life in the 21st century, including the ability to work in teams, to innovate, to solve complex problems, and to analyse and evaluate diverse information. Fourth, advances in technology have raised the possibility and challenge of fundamentally transforming assessment processes and information in the future.

The field of educational assessment is ill equipped to respond to these pressures for reform. It remains deeply divided into supposedly different assessment approaches and paradigms. The resulting dichotomies have become the default basis for conceptualising and describing

the field: quantitative versus qualitative; formative versus summative; norm-referenced versus criterion/standards-referenced; tests versus assessments; internal versus external; continuous versus terminal; measurement versus judgement; assessment *of* learning versus assessment *for* learning; and so on. In the absence of a unifying theory of assessment, the field remains fractured, divided into camps, and limited in its ability to respond to the opportunities and challenges it now faces.

Nevertheless, important progress has been made towards an alternative and more unified conceptualisation of educational assessment, including through the work of Pellegrino, Chudowsky and Glaser (2001) and Wilson (2005, 2009). The present review analyses recent pressures for assessment reform and their implications, reviews progress towards a more unified conceptualisation of the field, outlines a set of principles for a learning assessment 'system', and considers practical challenges in achieving assessment reform.

Demands for better information for decision-making

Over the past three decades, governments and education systems in many countries have given increased priority to improving the quality and equity of school education. In the United States of America, the report *A Nation at Risk* (National Commission on Excellence in Education, 1983) first raised public concerns about declining educational standards, the international competitiveness of American schools, and the large and continuing achievement gaps between 'white' and 'minority' students. More recently in Australia, similar concerns have been raised about educational performances in comparison with other countries in our region and continuing achievement gaps, particularly for disadvantaged students and Indigenous students. International concerns of these kinds have led to demands for better information about the outcomes of school education – the knowledge, skills and attitudes that students are developing through their schooling – as well as better information about trends in achievement levels over time and progress in closing achievement gaps.

These concerns have also led to more intense efforts to identify strategies for improving the quality and equity of educational provision. There has been growing global interest in the policies and practices of high-performing education systems, the characteristics of high-performing schools and the practices of high-performing teachers. The consequence has been a shift in emphasis from ensuring adequate role performance on the part of teachers, school leaders and system managers, to promoting specific, evidence-based improvement practices at all levels of an education system.

One general observation arising from these efforts has been the crucial importance of establishing appropriate starting points for action. It is well understood at the level of the classroom that successful learning is more likely when individual learners are given learning opportunities appropriate to their current levels of achievement and learning needs. But this principle of 'differentiation' appears to apply equally at the level of schools and education systems. Effective school improvement strategies appear to be different at different points in a school's improvement journey; and effective system improvement strategies appear to be different at different stages of a system's development (Mourshed, Chijioke & Barber, 2010). It follows that decisions about appropriate interventions and strategies, whether by classroom teachers, school leaders, system managers or governments, depend on good information about the prevailing circumstances. Educational assessment has a critical role to play in contributing to this information.

Effective decision-making also depends on an ability to monitor and evaluate the effectiveness of decisions and actions. Again, this is true at all levels of educational decision-making. Feedback is essential to making midstream adjustments and to evaluating the effectiveness and impact of interventions, programs and improvement strategies. Depending on the level of decision-making, the evidence required for this purpose includes information about the progress of individual learners, trends in average achievement levels over time and changes in achievement gaps. Again, educational assessment has an essential role to play in providing such evidence.

Demands for better evidence to inform decision-making have placed new expectations on educational assessment. Traditional expectations of assessment were based on a belief that the

role of teachers was to deliver the curriculum, the role of students was to learn, and the role of assessment was to establish how much of what teachers had taught, students had successfully learnt. Information about the success of student learning could be collected either mid-course and used to guide further teaching (that is, used formatively) or collected at the end of a course and used to judge overall student success (that is, used summatively). Many existing assessment processes, and much of the language of educational assessment, were developed for this traditional purpose of *judging* student success.

In contrast, the use of assessment to inform decision-making parallels the use of assessment in other professions such as medicine and psychology, where the purpose is not so much to judge as to *understand*. Professionals use assessments to better understand the presenting situation or problem, to identify starting points for action, to decide on appropriate evidence-based interventions, to monitor progress, and to evaluate the effectiveness of the decisions they make. As professional behaviours of this kind have become increasingly characteristic of all levels of educational practice, there has been a corresponding need for assessments that better inform decision-making processes.

Evolving understandings of human learning

A second set of pressures for assessment reform has arisen from research into learning itself. Our understanding of basic learning processes, impediments to learning and the conditions that support successful learning has continued to develop over recent decades, in part through research in disciplines such as cognitive science and neuroscience, as well as through professional experience and research into effective teaching and learning practices. The multidisciplinary study of learning, increasingly referred to as the 'science of learning', is providing insights that sometimes challenge long-held views about learning, and thus approaches to assessing and reporting learning.

One of these insights relates to the remarkable human capacity for learning. Research is demonstrating the brain's ability to learn throughout a person's lifespan, through the formation and strengthening of new pathways and neural networks (referred to as the brain's 'plasticity'). Learners may be at different points in their learning and may be progressing at different rates, but every individual seems capable of further learning if motivated and provided with appropriate learning opportunities. This research supports more positive and optimistic views of learning capacity than earlier beliefs that humans differed markedly in their innate capacity to learn, implying that assessments could be developed to identify these differing individual capacities. It now seems more useful to view learning as an ongoing, potentially lifelong process, and to view every learner as being on a path of learning, with the potential for further progress.

At the same time, research is showing that individuals of the same age can be at very different points in their learning. For example, in any given year of school, the most advanced learners in areas such as Reading and Mathematics can be as much as five or six years ahead of the least advanced learners. Coupled with the observation that successful learning is most likely when individuals are given learning opportunities appropriate to their readiness and needs, this finding underlines the importance of 'differentiated' teaching, which in turn depends on good information about where individuals are in their learning progress.

Much traditional practice in school education has been inconsistent with modern understandings of learning. Beliefs that individuals differ fundamentally in their capacity to learn has led in the past to efforts to identify these differing capacities (for example, through IQ testing) and to the creation of different pathways and streams (for example, academic, vocational) for students with differing perceived capacities. Rather than seeing all students as being on paths of learning and having the capacity for ongoing learning and, potentially, high achievement, traditional practices often lock students into streams which, in practice, place ceilings on how far they can progress (for example, by assigning them to low, intermediate and advanced Mathematics courses).

The ways in which schooling is structured and education is delivered can also be inconsistent with current understandings of learning. For example, despite the evidence that students of the

same age are at very different points in their learning, much teaching is focused on delivering the same year-level curriculum to all students in a class. Often this curriculum is divided into discrete units or courses. At the completion of each unit, students are assessed to establish how much of the content they have successfully learnt and their level of success is reported as a score, percentage or grade. Attention then shifts to the next unit of work (which may or may not build on the content of the previous unit) and the sequential process of teaching, learning and assessment re-commences. In contrast to a view of learning as a continuous process, the structure of schooling typically creates major disjunctures between units of learning, courses, years of schooling and phases of schooling. And, rather than being used to establish where students are in their long-term learning and to monitor progress over time, assessments are usually used to judge success on discrete bodies of taught content. As a consequence, assessment results usually fail to convey the progress that individuals are making in their learning (most clearly illustrated when a student is awarded the same grade year after year).

Current assessment and reporting practices are intimately tied to – and are usually designed to support – traditional approaches to schooling, including the K–12 assembly-line model, whole-class teaching, age-based curricula and the process of judging student success. Learning in the future is likely to be freed of many of these constraints as technology increasingly allows personalised learning anywhere at any time. For learning of this kind, it will be more appropriate that assessments provide information about where individuals are in their learning, what experiences and activities are likely to result in further learning, and what learning progress is being made over time.

Other research into learning has underlined the importance of knowledge transfer; that is, the application of understandings and skills to new, unseen contexts. Transfer is facilitated in part by deep understandings of the concepts, principles and key ideas of a learning area. In most areas of learning, extensive factual and procedural knowledge is essential to expert performance, but experts organise and apply their knowledge using a framework of concepts and principles (Bransford, Brown & Cocking, 2000). The implications for assessment are that greater priority needs to be given to the assessment of conceptual understandings, mental models and the ability to apply learning to real world situations.

Also contributing to research into learning have been studies of effective teaching practices. For example, studies of teaching have highlighted the crucial role that feedback plays in successful learning. To be effective, feedback must be timely and in a form that guides student action and builds confidence that further learning is possible. Findings from these studies, too, have implications for assessment and reporting. Feedback in the form of judgements of student success is much less helpful to learning than feedback that allows learners to understand where they are in their learning and so provides guidance on next steps (Hattie, 2003).

Finally, learning research is clarifying the important role of attitudes and self-belief in successful learning. Successful learners have strong beliefs in their own capacity to learn and a deep belief in the relationship between success and effort. They take a level of responsibility for their own learning (for example, identifying gaps in their knowledge and taking steps to address them) and monitor their own learning progress over time. The implications of these findings are that assessment processes must be designed to build and strengthen metacognitive skills. One of the most effective strategies for building learners' self-confidence is to assist them to see the progress they are making. As previously noted, current approaches to assessment and reporting often do not do this. When students receive the same letter grade (for example, a grade of 'B') year after year, they are provided with little sense of the progress they are actually making. Worse, this practice can reinforce some students' negative views of their learning capacity (for example, that they are a 'D' student).

An emphasis on broader life skills and attributes

A third set of pressures for assessment reform have their origins in concerns about the extent to which current school curricula are developing the knowledge and skills necessary for work and life in the 21st century. Employers, in particular, have expressed concerns about how well prepared school leavers are for today's workplaces and have emphasised the need for employees who can work collaboratively in teams, use technology effectively and create new solutions

to problems. These concerns have led to new levels of effort to identify, develop and assess a broader range of skills and attributes within the school curriculum.

A prominent example of such an effort is the international collaboration known as the Assessment and Teaching of 21st Century Skills (Griffin, McGaw & Care, 2012). This collaboration has identified four broad categories of skills: ways of thinking (including creativity, critical thinking, problem-solving, decision-making and learning); ways of working (including communication and collaboration); tools for working (including information and communications technology and information literacy); and skills for living in the world (including citizenship, life and career, and personal and social responsibility).

Inherent in initiatives of this kind is the view that assessment has a pivotal role to play in focusing the attention of schools and school systems on the development of broader life skills and attributes. As long as assessment and reporting processes retain their focus on the mastery of traditional school subjects, this focus will continue to drive classroom teaching and learning. There is also growing recognition that traditional assessment methods, developed to judge student success on defined bodies of curriculum content, are inadequate for assessing and monitoring attributes and dispositions that develop incrementally over extended periods of time.

The transformational potential of new technologies

A final set of pressures for assessment reform is coming from rapid advances in technology. As new technologies are more widely used for the delivery of educational courses and as an increasing proportion of learning occurs online outside formal educational arrangements, it is clear that assessment processes will also become increasingly technology-based. The greater use of technology introduces the possibility of significantly transforming assessment processes. For example, assessments can be based on student responses during online learning activities; online tasks can be automatically selected, based on students' demonstrated levels of performance (as in computer adaptive testing); and more intelligent assessments can be developed, for example to formulate and test automatically hypotheses about the nature of students' errors or misunderstandings. Assessments of the future are likely to make greater use of simulated learning environments, including three-dimensional virtual environments in which students can manipulate variables and perform activities that may be difficult or impossible to create in normal classrooms.

Important work has been done to develop new technology-based forms of digital assessment, including for national assessments of ICT literacy (Ainley, Fraillon, Gebhardt & Schulz, 2012) and international assessments of digital reading (Thomson & De Bortoli, 2012). However, many early efforts to use technology for assessment have been limited to the delivery of traditional test items on screen or to the development of collections of online assessment tasks as resources for teachers. These are relatively pedestrian uses of technology and are likely to be superseded in the future by much more powerful forms of assessment.

Towards a unified theory of assessment

As already noted, the field of educational assessment is fragmented, limiting its ability to respond to emerging pressures and developments of these kinds. The field has been divided into perceived 'multiple purposes' of assessment (for example, formative, summative, diagnostic, screening and large-scale surveys), which are often assumed to require entirely different approaches and are sometimes characterised as being in opposition (for example, formative versus summative assessments). The field has also been divided into varying assessment 'methods' (for example, school-based versus external, tests versus assessments, authentic versus devised). Individual methods have attracted their proponents, who often view a specific method as inherently superior to others. The consequence of this fragmentation is a field characterised by contrasting philosophies, methods and purposes, and an often unproductive discourse.

However, in recent years, progress has been made towards a more unified view of educational assessment through a simple unifying principle; namely, that *the fundamental purpose of*

assessment is to establish where learners are in their learning at the time of assessment. Rather than beginning from the position that educational assessments have disparate purposes, this principle recognises that all assessments have the same fundamental purpose and that there are then varying uses to which assessment information can legitimately be put. Rather than beginning with a belief that some methods of assessment are intrinsically superior to others, this principle recognises that the most appropriate method(s) in any particular assessment context are those that provide the most practicable, valid and reliable information about where learners are in their learning.

Underpinning this unifying principle is a view of learning as a process that occurs over multiple years, and is potentially lifelong. Assessment, rather than being simply a process of judging how well students have learnt what they have just been taught, is conceptualised as the process of establishing where students are in their long-term learning and what progress they are making over time, usually in terms of their developing knowledge, skills and understandings.

Information about where students are in their learning can be used in many different ways, including to identify starting points for teaching, to diagnose errors and misunderstandings, to monitor trends in average achievement levels over time, to select students for entry into courses, to evaluate the effectiveness of teaching interventions, and to benchmark achievement levels against international standards. All these *uses* depend on good assessment information about where students are in one or more aspects of their learning.

The question of where students are in their learning can be asked and answered at different levels of detail. For example, OECD PISA assessments gather information about the average reading literacy levels of Australian 15-year-olds. But by zooming in to finer levels of detail, questions also can be asked about the average reading literacy levels of Indigenous Australian 15-year-olds, average reading literacy levels in a particular school, or even about the reading literacy level of a particular 15-year-old student. Although different strategies (for example, sampling) may be required to collect assessment information at these different levels of detail, the fundamental purpose – to establish where 15-year-old students are in their reading literacy development – remains the same.

The question of where students are in their learning can also be asked and answered at different levels of detail in relation to the learning domain. For example, while OECD PISA assesses the domain of reading literacy, at a finer level of detail PISA assesses three Reading sub-domains (accessing/retrieving, integrating/interpreting and reflecting/evaluating) and reports separately on student achievement in each sub-domain. By looking in increasing detail at where students are in their learning, assessments can provide increasingly diagnostic information about areas of strength and weakness. Forster (2009) refers to the level of detail that an assessment instrument is designed to inspect as its diagnostic 'power'. Diagnostic assessments require information about relatively detailed aspects of learning, but have the same fundamental assessment purpose: to establish and understand where learners are in their learning.

Thus the conceptualisation of educational assessment as the process of establishing where learners are in their learning at the time of assessment can be applied equally to a national sample of students participating in an international survey and to the diagnosis of an individual learner's specific difficulties within a relatively narrow aspect of their learning.

Under this more unified conceptualisation of assessment, many of the popular dichotomies and distinctions of the field become less relevant. When assessments are used to establish where students are in their learning, the same assessments can be used to identify starting points for further teaching and learning and/or to review the progress that individuals have made since some previous assessment (an assessment of the learning that has occurred). For diagnostic and teaching purposes, teachers may sometimes find it useful to examine in more detail where students are in specific aspects of their learning, but assessments *of* learning and assessments *for* learning need not be fundamentally different kinds of assessments; they are more helpfully conceptualised as different and complementary uses of information about student learning.

Similarly, it has been common to represent 'norm-referenced' assessments and 'standards-referenced' assessments as being different and often diametrically opposed forms of assessment.

But under a more unified view of assessment, norm-referencing and standards-referencing (or criterion-referencing) are more appropriately conceptualised as different ways of *interpreting* information about where students are in their learning. Once information is available about where a student is in his or her learning, that information can be interpreted in a variety of ways, including in terms of the kinds of knowledge, skills and understandings that the student now demonstrates (criterion- or standards-referencing); by reference to the performances of other students of the same age or year level (norm-referencing); by reference to the same student's performance on some previous occasion; or by reference to a performance target or expectation that may have been set (for example, the standard expected of students by the end of Year 5). Once it is recognised that the fundamental purpose of assessment is to establish where students are in their learning (that is, what they know, understand and can do), many traditional assessment distinctions become unnecessary and unhelpful.

Design principles for a Learning Assessment System

This conceptualisation of assessment as the process of establishing and understanding where learners are in their learning has major implications for assessment design and practice in classrooms, schools and school systems. Under this conceptualisation, all steps in an assessment process are designed to address this essential purpose, imposing demands not generally satisfied by assessments conducted merely to judge student success on a taught body of curriculum content. The steps in the process are interdependent and constitute a coherent 'system', referred to in this review as a Learning Assessment System.

Underpinning a Learning Assessment System is a set of five design principles. These are briefly outlined here and elaborated in Section 3.

Principle 1: *Assessments should be guided by, and address, an empirically based understanding of the relevant learning domain.*

The monitoring of learning progress requires deep familiarity with the terrain through which learners are progressing. Skilled monitoring depends on knowledge of how learning typically progresses within the domain – for example, an understanding of the prerequisite skills and knowledge for successful further learning and an awareness of common misunderstandings, errors and obstacles to learning progress. Deep knowledge of this kind is based on more than familiarity with an intended curriculum. It depends on accumulated professional experience and research into the nature of learning within the relevant domain.

Principle 2: *Assessment methods should be selected for their ability to provide useful information about where students are in their learning within the domain.*

The primary consideration in choosing an assessment method should be the method's ability to provide valid information about the domain in question. Different assessment methods, including electronic assessments, paper and pen tasks, student performances, research projects, products of art and technology and portfolios of student work are likely to be valid for different kinds of learning. For example, detailed diagnostic testing may be necessary to understand the exact nature of a learner's difficulties or a focused one-on-one conversation may be necessary to explore and understand a particular student misconception. Once a general method of assessment has been chosen, specific assessment activities or 'tasks' are required. In developing assessment tasks, consideration needs to be given to a range of other criteria, including reliability, objectivity, inclusivity and feasibility.

Principle 3: *Responses to, or performances on, assessment tasks should be recorded using one or more task 'rubrics'.*

Each task rubric consists of two or more ordered levels or response to a task. Generally, responses/performances are recorded using a single rubric for a task, but for large and complex tasks, several rubrics might be used to assess and record different aspects of performance (sometimes referred to as analytical scoring). Importantly, task rubrics provide the direct substantive link

to the larger learning domain. Through their ordered levels of response/performance, they operationalise what it means to make progress within the domain.

Principle 4: *Available assessment evidence should be used to draw a conclusion about where learners are in their progress within the learning domain.*
This conclusion, which is the purpose of the assessment process, always involves an interpretation of the available assessment evidence. It is an inference based on recorded task performances. In some contexts, the main purpose of an assessment process may be to establish whether or not (or how well) a person can perform a specific task. In other words, the interest is in the task itself. However, in educational contexts, individual assessment tasks are very rarely of intrinsic interest. They are merely convenient and interchangeable vehicles for gathering evidence and drawing conclusions about where learners are in their learning within the domain of interest.

Principle 5: *Feedback and reports of assessments should show where learners are in their learning at the time of assessment and, ideally, what progress they have made over time.*
The focus is on the stage learners have reached in their learning, usually interpreted in terms of the kinds of knowledge, skills and understandings that they now demonstrate. Feedback and reporting of this kind are based on a conception of learning as an ongoing, long-term process. Such reports are likely to be quite different in appearance from reports based on judgements of how well students have learnt what they have been taught.

This review paper argues that reform to meet the demands now being placed on educational assessment requires the adoption and implementation of a coherent assessment 'system' based on a set of principles of this kind.

Structure of this review

Section 1 introduces the topic of this review, briefly outlines some current pressures for assessment reform, introduces the concept of a Learning Assessment System designed to establish where learners are in their progress within an empirically mapped domain of learning, and sketches a set of design principles for such a system.

Section 2 considers pressures for assessment reform in greater detail and the implications of these pressures for educational assessment practice. The conclusion of this section is that current pressures for reform point to the need for assessments based on a view of learning as a personal, long-term process. Traditional methods designed to judge student success on defined bodies of curriculum content are of limited value outside traditional educational delivery structures, when the focus is on deep understandings and life skills that develop only over extended periods of time, given the currently observed variability in students' levels and rates of progress, and in the context of demands for better information to guide all levels of educational decision-making.

Section 3 describes the elements of, and design principles that underpin, a Learning Assessment System. The elements are described in some detail: an empirically based learning domain; a domain-relevant assessment method or methods; task rubrics for recording students' responses to, or performances on, assessment tasks; and a process for bringing together task-based evidence to draw conclusions about where students are in their learning within the domain.

Section 4 considers some practical challenges in implementing learning assessment systems of this kind. These challenges include changing widely held perceptions that educational assessment is fundamentally a process of judging student success; the development of deep understandings of how learning occurs within specific learning domains as a basis for establishing where learners are in their learning; the promotion of more coherent systems of assessment across a range of educational contexts; and the promotion of higher levels of assessment literacy across the profession.

section 2

The need for assessment reform

Over recent decades, higher expectations have been placed on the assessment of student learning. There has been growing interest on the part of governments and education systems in understanding how national and state/territory achievement levels compare with international standards, and in understanding what high-performing education systems do to achieve high levels of excellence and equity in school education. There has been growing interest in the practices of schools that make rapid improvements in their student achievement levels or that produce results better than expected from their student intakes. And there has been growing interest in how highly effective teachers use assessment as part of their teaching to promote improved student learning (Black & Wiliam, 1998; Wiliam & Thompson, 2007).

In parallel with this growing interest in the use of assessment to better understand effective practices and to inform educational decision-making have been efforts to broaden assessment methods beyond traditional approaches, particularly paper and pen tests and examinations, and to ensure that assessment and reporting processes are better aligned with current understandings of learning. Calls for schools to develop and assess a wider range of life skills and attributes have introduced a new set of assessment challenges. And, in parallel with all this, advances in technology are raising the possibility of fundamentally transforming learning assessments and information in the future. This section explores some of these pressures in more detail and considers their implications for practice.

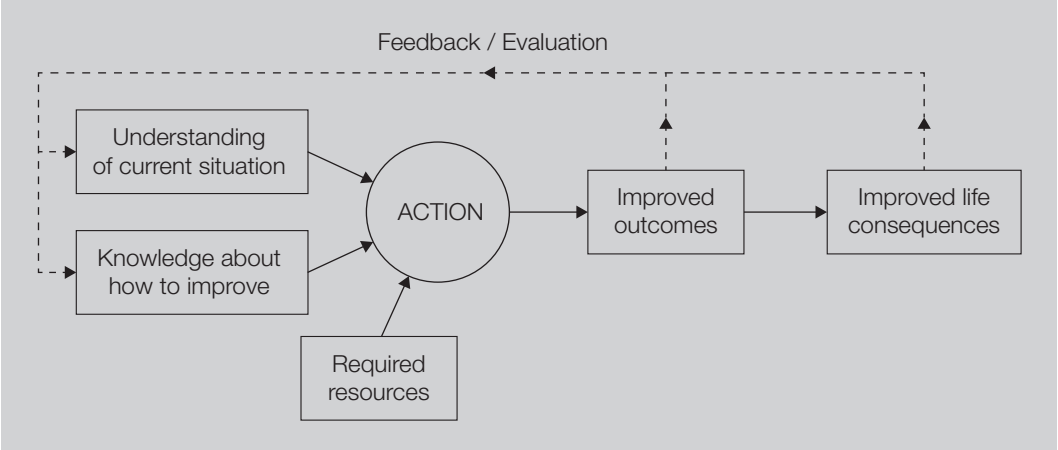
Demands for better information for decision-making

Better information is now being sought to inform all levels of educational decision-making, including but not limited to, classroom teachers, early childhood specialists, teachers in vocational and higher education settings, others who work in professional support roles, institutional leaders, staff of district, regional and central offices, leaders of education systems, and national and state governments. In all areas of educational decision-making, assessments of student learning are increasingly being sought to assist in:

- clarifying starting points for action
- investigating details of student learning and performance
- monitoring improvements and evaluating educational interventions
- motivating effort and encouraging self-monitoring
- providing feedback to guide future action.

The crucial role of assessment in educational decision-making is illustrated in Figure 2.1. This diagram is referred to as an educational decision-making 'loop' because it represents an iterative process through which feedback on past decisions and actions informs future practice.

Figure 2.1: Educational decision-making loop



Clarifying starting points for action

Educational work of all kinds involves making decisions and taking action for the purposes of improving learning outcomes and delivering long-term benefits for learners. Professional *action* is at the centre of the decision-making loop. Effective action depends on a good understanding of the presenting situation combined with deep professional knowledge about strategies that are likely to be effective in delivering improved outcomes (see the inputs on the left of Figure 2.1).

Assessment plays an essential role in clarifying starting points for action. This is a feature of professional work in all fields. Professionals such as architects, engineers, psychologists and medical practitioners do not commence action without first gathering evidence about the situation confronting them. This data-gathering process often entails detailed investigation and testing. Solutions, interventions and treatments are then tailored to the presenting situation or problem, with a view to achieving a desired outcome. This feature of professional work distinguishes it from other kinds of work that require only the routine implementation of pre-prepared, one-size-fits-all solutions.

The decision-making loop in Figure 2.1 is relevant at all levels of educational action. In classrooms, teachers require good understandings of where students are in their learning (that is, their current levels of knowledge, skill and understanding) to design learning opportunities appropriate to individuals' or groups' levels of readiness and learning needs. Educational psychologist David Ausubel identified this feature of effective teaching more that 40 years ago:

> *If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly.*

(Ausubel, 1968, p. vi)

The process of 'ascertaining' what the learner already knows is the process of assessment, which in classrooms can be undertaken in a variety of ways. Assessments to understand where learners are in their learning and to identify starting points for teaching require active investigation on the part of the teacher. This process of investigating student learning to identify appropriate teaching strategies is considerably more demanding, but more consistent with the nature of professional work, than the delivery of the same grade-based curriculum in the same way to all students in a class.

And it is not only teachers who need to clarify starting points for action. Leaders of educational institutions and school systems also require good information about where students are in their

learning as a basis for informed action. School leaders are in stronger positions to intervene and to take appropriate action to improve student outcomes if they have reliable information about current levels of student achievement (for example, information about literacy and numeracy levels in a school and, perhaps, how those levels compare with year-level expectations and levels in other, similar, schools).

Leaders of education systems are in stronger positions to take action if they have reliable information about how students are performing across an entire system; for example, information about levels of mathematics and science learning and, perhaps, how those levels compare with levels in other education systems and countries. Information of this kind can be helpful in identifying and clarifying problems or challenges that may not previously have been fully appreciated. For example, information from the OECD's Programme for International Student Assessment (PISA) in 2000 alerted German authorities to the relatively poor performances of German 15-year-olds by international standards. Similar information from the National Assessment Program, Literacy and Numeracy (NAPLAN) and the IEA's Trends in International Mathematics and Science Study (TIMSS) in Queensland in 2008 raised concerns about levels of literacy, numeracy and science learning in primary schools in that state. In both cases, by clarifying existing achievement levels, assessments informed system action.

Investigating details of student learning

The process of clarifying where learners are in their learning and identifying starting points for action sometimes requires more in-depth explorations of the specifics of student learning. At the classroom level, effective teaching interventions occasionally require more detailed investigations of students' misunderstandings and difficulties, perhaps through focused teacher questioning or diagnostic testing. Such processes could, for example, involve the formulation of hypotheses about the nature of an individual's difficulties or misconceptions, followed by the collection of evidence to test those hypotheses.

At the level of a school, the instructional leadership required to achieve improved student outcomes usually requires the identification of specific areas in which the school is performing well or poorly. This may include a detailed analysis of performances across year levels, classrooms or identified subgroups of the student population. More detailed information of this kind enables school leaders to target resources and action in the areas of greatest need within their school.

Similarly, at the level of an education system or nation, assessments are capable of identifying specific areas of learning or subgroups of the student population requiring special attention. For example, assessments conducted as part of the 2007 Trends in International Mathematics and Science Study (TIMSS) revealed that Australian Year 8 students performed at relatively low levels in algebra, suggesting that this may be an area requiring greater attention (Thomson, Wernert, Underwood & Nicholas, 2008). Assessments conducted as part of the 2009 Programme for International Student Assessment (PISA) showed that the overall decline in reading levels of Australian 15-year-olds since 2000 was attributable in part to a decline within a specific subgroup (high-performing students), again suggesting an issue that may require greater attention and action.

Monitoring outcomes and evaluating effectiveness

Because the purpose of educational work is to bring about improved learning outcomes for students, reliable measures of outcomes (see right of Figure 2.1) are essential to the evaluation of educational practice. The term 'outcomes' is used very broadly here to include whatever student knowledge, skills, understandings, attitudes, values or dispositions educational practitioners are striving to develop. The question to be asked of all educational practice is: Is it contributing to improved outcomes?

Assessments of student progress allow teachers to track and address the changing needs of individual learners over extended periods of time. They also allow teachers to make judgements about the adequacy of student growth, to identify students who are slipping behind in their

learning and to intervene in an effort to put individuals back on successful learning trajectories. Teachers also require measures of improvement to evaluate the effectiveness of their teaching strategies and interventions. Assessments of student progress can provide such data.

Leaders of educational institutions and systems require information about how student achievement levels are changing over time to inform decision-making and to evaluate the effectiveness and impact of educational programs and initiatives. For some purposes, information is required about the *gains* students make across the years of school (for example, how do average reading gains from Year 3 to Year 5 in this school compare with average reading gains in similar schools?). For other purposes, information is required about *trends* in performance levels over time (for example, have the mathematics levels of 15-year-olds in this country improved, declined or stayed the same over the past decade?). Information of this kind, derived from reliable assessments, is essential to answering questions such as the following:

- Did this government initiative improve mathematics learning?
- Have average reading levels in this system changed over time?
- Are achievement gaps between identified student subgroups closing?
- Which of these programs was more effective in promoting student learning?

Providing feedback to inform future action

Feedback completes the decision-making loop in Figure 2.1 by refreshing knowledge about the current situation (for example, student achievement levels) and by contributing to accumulating professional knowledge about effective courses of action. In these ways, feedback informs and provides starting points for future action in a cycle of potentially ongoing improvements.

At the classroom level, research shows that the provision of feedback to students is one of the most effective strategies for promoting further learning (Hattie, 2003). To be most effective, feedback must be timely, must be in a form that encourages effort and that allows learners to see the progress they are making. Feedback must identify clear actions that individuals can take to make further learning progress.

In summary

Professional practice at all levels of school education requires dependable information to establish starting points for action, to understand areas of strength and weakness, to monitor improvements over time, to evaluate the effectiveness of programs and interventions, to motivate improvement and encourage self-monitoring, and to provide feedback to guide further action.

Evolving understandings of human learning

A second general pressure for assessment reform has emerged from advances in our understanding of basic learning processes and the conditions that support successful learning. In recent decades, learning research in a number of disciplines, including neuroscience and cognitive science, has enhanced our understanding of human learning. Some of this research is providing deeper insights into the nature of learning within specific domains. Studies of effective pedagogical practices also are contributing to our understanding of the conditions for successful learning. These evolving research-based understandings are sometimes challenging traditional approaches to educational delivery, including approaches to assessing and monitoring student learning.

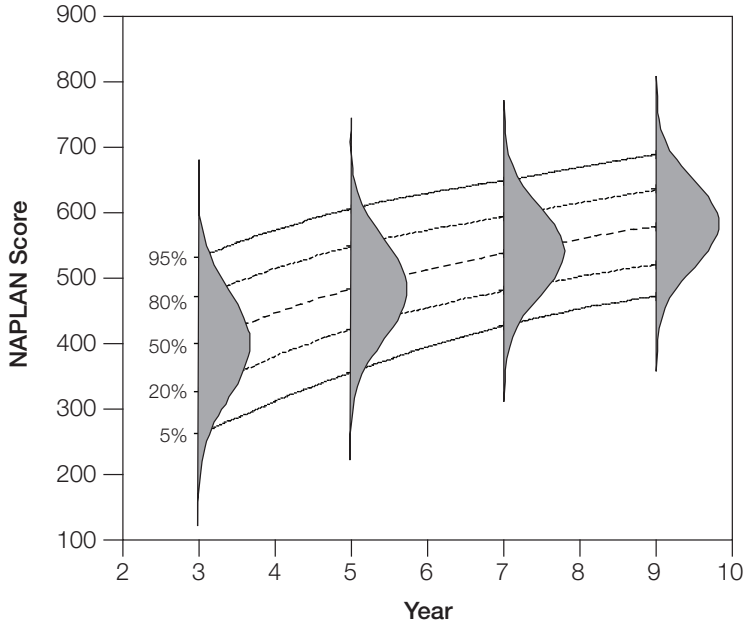
Variability in student learning

Evidence from educational assessment programs designed to establish where students are in their long-term learning is revealing enormous variability in learners' levels of achievement. Research suggests that this variability is due in part to individuals' past experiences and to the environments to which they have been exposed. For example, research in neuroscience

shows that the brain is shaped by experience. Although basic brain structures are common across individuals, the human brain adapts to its environment and develops in response to the experiences of the individual. These differences commence before birth with prenatal conditions influencing brain development. In the early years of life, an individual's experiences influence which neuronal connections are established and which are weakened and removed. At older ages, significant and permanent changes in brain structures and functioning occur as time is spent learning and practising particular procedures and skills. And the ability of the brain to build, reorganise and dismantle networks of neurons – known as the brain's 'plasticity' – continues well into old age. In short, neuroscience reveals learning to be a highly personal process, with the consequence that individuals of the same age can be at very different points in their learning and development.

By the time children commence school, they have markedly different levels of cognitive, language, emotional, social and psychomotor development. Importantly, significant differences in these areas of learning and development persist throughout the years of school, and beyond. As an example, the observed variability in Australian students' reading levels is shown in Figure 2.2. This figure, developed by the author based on the responses of all Year 3, 5, 7 and 9 students to reading tests conducted as part of the National Assessment Program, Literacy and Numeracy (NAPLAN) in 2008, enables the cross-year comparison of students' reading levels.

Figure 2.2: Distributions of students' reading abilities (Years 3, 5, 7 and 9, Australia, 2008)



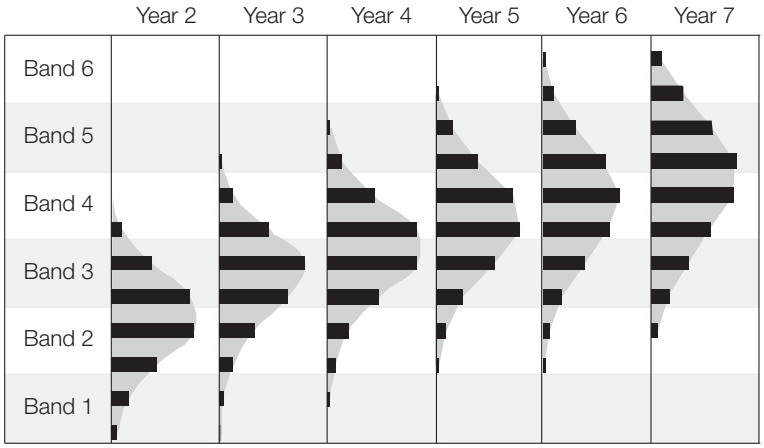
(Masters, data file)

Figure 2.2 shows a wide spread of reading levels in each of these four years of school. Based on average reading gains between Year 3 and Year 5, the highest-performing 10 per cent of students in each year of primary school are about five years ahead of the lowest-performing 10 per cent of students in that year. The consequence of this variation is that the highest-performing 20 per cent of Year 3 students are better readers than the lowest-performing 5 per cent of Year 9 students. An identical conclusion was reached in an earlier national sample survey of Australian primary students (Masters & Forster, 1997a). Figure 2.2 also shows that, although this spread decreases slightly by the secondary years, so does the average rate of growth, with the consequence that the highest-performing 10 per cent of Year 9 students continue to be five or six years ahead of the lowest-performing 10 per cent of students.

Very similar variability has been observed internationally in school mathematics achievement. For example, assessments of mathematics learning in the United Kingdom show that, by the

end of primary school, students differ by at least six years of school (Harlen, 1997; Wiliam, 2007). But, unlike assessments of reading, assessments of mathematics learning sometimes show *increasing* variability in students' achievement levels across the years of school.

Figure 2.3: Distributions of students' mathematics achievements (Years 2–7, USA, 2003)



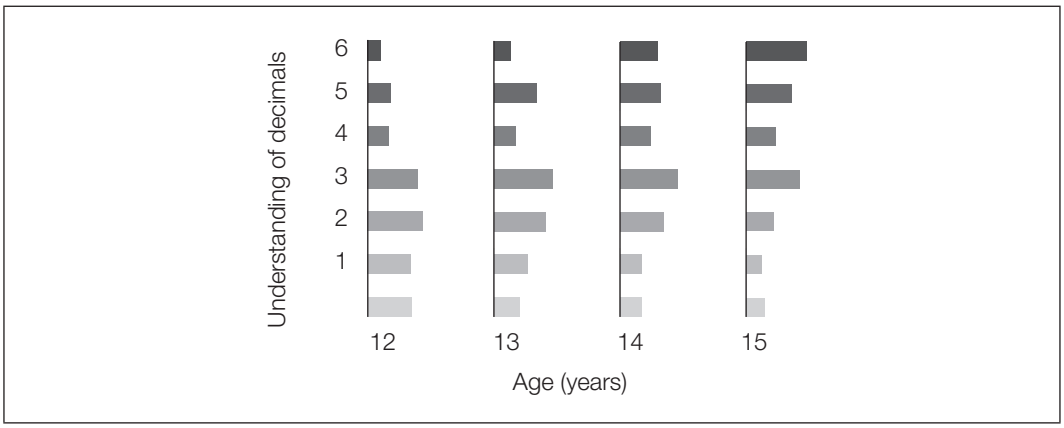
(Masters, data file)

Figure 2.3, also developed by the author, is based on assessments of more than a quarter of a million students in the United States of America (Hauser, 2003). It shows significant overlap and increasing spread of achievement distributions between the second and seventh grades. It might be anticipated that the longer students are in school, the more similar their levels of achievement would become. In tests of Mathematics, at least, it appears that the opposite may be true. Closer inspection of the data in Figure 2.3 suggests that the reason for this increasing spread is that, while the highest achieving students appear to make steady progress, the tail of the distribution falls increasingly far behind with each year of school. Increasing variability in mathematics achievement across the years of school has also been observed in a range of standardised mathematics tests internationally (Wiliam, 2007).

Both Figures 2.2 and 2.3 also show slowing rates of reading and mathematics progress into the secondary years of school. This is a commonly observed phenomenon; progress in many learning areas appears to occur most rapidly in the early years of school. Not surprisingly, given the slower average rates of progress in the post-primary years, significant variability and overlap of achievement distributions are evident throughout the secondary school years.

A third example, Figure 2.4, shows the distributions of UK students' understandings of decimals between 12 and 15 years of age. These assessments were made as part of a national sample survey conducted in the United Kingdom (Wiliam, 2007).

Figure 2.4: Distributions of students' understandings of decimals (ages 12–15, UK)



(Wiliam, 2007, p. 248)

As expected, the percentage of students in the highest achievement level (Level 6) increases with age, and the percentage in the lowest level decreases with age. But the striking feature of these data is the variability of students' understandings of decimals at each age and the very small amount of average improvement in these understandings between 12 and 15 years of age, leading Dylan Wiliam to observe that:

> *Attainment is only loosely related to age.*

(Wiliam, 2007, p. 248)

This observed variability in student learning has important implications for teaching. Clearly, any attempt to infer students' learning needs, or to base starting points for teaching, on age or year level alone, would run the risk of being very wide of the mark for some individuals. A considerable body of research shows that optimal learning occurs when learners are presented with challenges just beyond their current level of attainment. This is what Vygotsky (1978) referred to as the 'zone of proximal development' – the region of 'just manageable difficulties' where students can succeed, but often only with the support of others, for example through 'scaffolding' activities. Highly effective teachers go to the trouble of understanding where individuals are in their learning and use this knowledge to identify appropriate starting points for teaching, thus maximising the likelihood of successful learning. In other words, effective teachers undertake assessments of where learners are in their learning before they start teaching.

> *There is a good deal of evidence that learning is enhanced when teachers pay attention to the knowledge and beliefs that learners bring to the learning task and use this knowledge as a starting point for new instruction.*

(Bransford, Brown & Cocking, 2000, p. 11)

However, in practice, it is not uncommon for teachers to begin teaching a class with only very limited understandings of what students in that class already know and can do. In these classrooms, rather than identifying appropriate starting points for teaching, teachers take their cue from the year level of the class and the associated year level curriculum ('I'm a Year 8 teacher, this is a Year 8 class, so I'll teach the Year 8 curriculum'). By limiting their efforts to covering the curriculum for the year level, these teachers tend to teach to the middle of the class at the expense of students at both extremes. Students who are not yet ready for that curriculum often fail to engage and thus fall further behind. (The observation that students in the tail of the mathematics distribution in Figure 2.3 fall further behind with each year of school may well be the result of exposure to year-level mathematics curricula that are increasingly distant from, and irrelevant to, their actual learning needs.) At the same time, more advanced students in a class often remain unchallenged and make limited progress (Griffin, 2013). When teachers focus their efforts only on delivering the curriculum for the year level, they often consider their job done when students finish assigned class work. Rather than extending more able students who finish class work early, teachers sometimes give these students 'free time'.

One-size-fits-all approaches to classroom teaching are based on the assumption that students of the same age and year level are more or less equally ready for the same curriculum – an assumption that clearly is not supported by the research evidence. Darling-Hammond (2004) refers to this assumption as the 'industrial, assembly line model of schooling' under which students are grouped according to age and move along a 'conveyor belt' from one year of school to the next. At each stage along this 'assembly line', students are delivered the curriculum considered appropriate for their age. In other words, starting points for teaching are based not on what learners bring to the classroom, but on external beliefs about what students of a particular age should be learning.

When teaching is focused only or primarily on the delivery of the relevant year level curriculum, there may be little interest in what has gone before; prior curricula are assumed to have been covered and learnt. Information about an individual's learning history that may have been accumulated, and that could be available at the start of a new school year or at the start of a new phase of schooling, may be seen as largely irrelevant to the task at hand.

> *In elementary schools, children move from one teacher to the next every year. Every year we trash a year's worth of relationships built between children and their teacher, and we throw away all the knowledge the teacher has gained about what each child needs and can do. Each year, we tell every child and teacher to start over again.*

(Marshak, 2003, p. 230)

For teachers, there are obvious practical challenges in identifying where each individual is in his or her learning, and in continually monitoring that student's progress over time. Nevertheless, this is exactly what effective teaching requires. In the future, technology may provide greater assistance to teachers in assessing, diagnosing and tracking individual learning progress. With better information of this kind, teachers will be in better positions to identify starting points for action and to decide on appropriate teaching interventions for their students.

> *In an ideal world, the teacher would have precise and current knowledge of each student's starting points and also of what assistance each student requires to move to the next level.*

(Fullan, Hill & Crévola, 2006, p. 34)

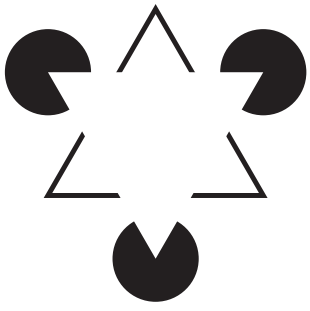
Learning as constructing meaning

Other learning research is revealing the active, constructive nature of human learning. It is now well understood that people learn by building representations of the world around them and by then attempting to interpret new information in terms of those existing understandings.

At the most basic level, research in neuroscience is showing that learning involves the building of networks of interconnected neurons. Each neuron can be connected to as many as several thousand other neurons organised into extensive networks, often across a number of regions of the brain. Learning involves the creation and strengthening of connections (synapses) between neurons and the strengthening of pathways in the brain by building insulation (myelin) around axons, enabling the faster transmission of electrical signals between connected cells. Far from being a passive process of storing incoming information in 'unused' parts of the brain, learning is an *active* process of relating incoming information to what is already known, through the construction and reconstruction of neuronal networks.

A simple illustration of the brain's attempt to make sense of incoming information is provided by the classic Kanizsa Triangle illusion (Figure 2.5).

Figure 2.5: Visual input leading the brain to 'see' an inverted white triangle



(Kanizsa, 1955)

When presented with the visual input in Figure 2.5, the brain 'sees' an inverted white triangle (which often appears brighter than the white background), overlaid on a second triangle, visualising boundaries where they do not actually exist. This illusion is an illustration of the brain attempting to interpret input in terms of pre-existing expectations.

The brain continually builds its understanding of the world by integrating new information with existing knowledge and understandings. This process occurs as the brain establishes,

strengthens, reorganises and dismantles networks of neurons. One of the most powerful forms of learning occurs when connections are built that produce a new insight or enable the learner to make sense of information in a new way. This is the 'ah-ha' moment – the flash of insight that provides a new level of meaning. The new neuronal connections made in this process provide the individual with a deeper understanding and an enhanced way of making sense of the world.

> *Of the many triggers that motivate people to learn … one of the most (if not the most) powerful is the illumination which comes from understanding. The brain responds well to this, which happens for instance during the 'eureka' moment, when the brain suddenly makes connections and sees patterns between the available information. It is the most intense pleasure the brain can experience, at least in a learning context.*

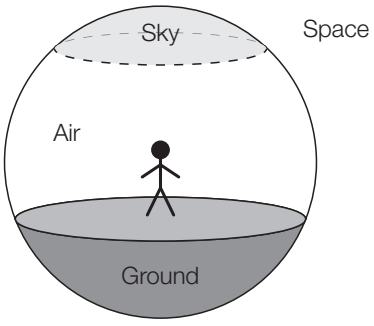
(Organisation for Economic Co-operation and Development, 2007, pp. 71–73)

Research in cognitive psychology and education continues to reveal that learning is a personal, active, constructive, rational process. Early childhood researchers have demonstrated that even very young infants actively work to make sense of the world around them. For example, if a child in the first year of life is shown two objects being placed behind a screen and the screen is then removed to reveal only one object, the child will look longer at that one object than when the process is repeated and the screen is removed to reveal two objects as expected – suggesting that the child is attempting to make sense of what they have just observed.

Through research in psychology and education, we now know that young children are capable of much more sophisticated reasoning than was once thought possible. Young children construct relatively sophisticated understandings of the world around them, and the challenge for them when they commence kindergarten or school is to connect what they are seeing and hearing to what they already know.

An instructive example of this comes from the work of Nussbaum (1979) who asked very young Israeli children to draw pictures of the Earth. He found that children drew different kinds of pictures, which he then categorised. Figure 2.6 shows one category of picture drawn in Nussbaum's study. It appears to demonstrate a child's attempt to integrate their experiential knowledge (that the Earth is flat) with what they have been told (that the Earth is round).

Figure 2.6: One kind of picture young children draw of the Earth



(Nussbaum, 1979, p. 88)

There is a large body of educational research illustrating the varying conceptions and mental models that students develop in relation to particular aspects of their learning. This research is a reminder that learning is very rarely a matter of adding knowledge to an empty vessel. Even at birth the brain is not a void. Learning, at all stages of life, is an active process of building connections and interpreting new information in terms of existing knowledge and beliefs.

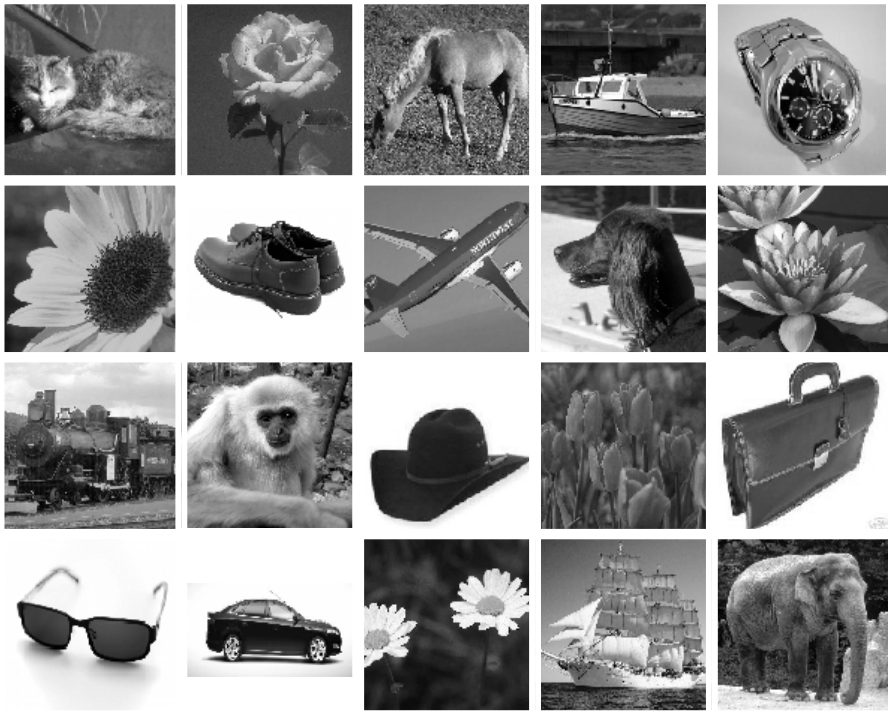
> *One major tenet of cognitive theory is that learners actively construct their understanding by trying to connect new information with their prior knowledge.*

(Pellegrino, Chudowsky & Glaser, 2001, p. 62)

Studies of novices and experts in a range of areas of expertise show that experts have deep and extensive knowledge of their subject matter. But perhaps equally importantly, experts develop ways of organising and making sense of this knowledge (referred to as 'schemas'). The extensive knowledge of experts is organised around important concepts, principles and big ideas in their areas of expertise, giving them deep understandings of their subjects. As a result, experts are able to see patterns, relationships and discrepancies in information that are not visible to novices. Their deep understandings also enhance their ability to select and remember relevant information. For example, chess masters are better able to recognise and memorise meaningful configurations of chess pieces on a board than novice players (Bransford, Brown & Cocking, 2000).

The usefulness of schemas for making sense of information can be simply illustrated using the images in Figure 2.7.

Figure 2.7: Twenty images used in a memory task



(Masters, 2009)

When adults are shown these 20 images and given a limited time to memorise as many of the objects as possible, they begin by noting that there are different categories of images (flowers, animals, modes of transport, accessories) and they use these categories as an aid to memorising the 20 objects. In other words, because adults are highly familiar with the 20 objects pictured, they are able to use their prior understandings to see patterns in the provided information. This helps them to memorise. In a similar way, experts in other fields are able to use the schemas they have developed to make sense of, and to remember, information in ways that are not usually available to novices.

The ability of experts to make sense of observations and information in terms of underlying principles, concepts or key ideas allows them to apply what they know to new contexts and problems and to transfer their learning to new situations. The knowledge of experts is described as 'conditionalised', meaning that they know when, where and how to apply their expert knowledge (Bransford, Brown & Cocking, 2000). Deep understandings of this kind are developed only over extended time periods through opportunities to observe the same ideas at work in multiple examples and in a range of applied contexts.

Consistent with findings in cognitive psychology that the development of expertise involves not only the development of extensive factual and procedural knowledge, but also the development of

schemas for organising and making sense of that knowledge, a considerable body of educational research has focused on students' conceptual understandings, which include their mental models of subject matter and the way these models change as students develop more sophisticated understandings of what they are learning.

Research evidence highlighting the importance of deep understandings to learning and to the development of expertise has obvious implications for teaching. Because learners construct their own mental models and understandings of what they are taught, and because these models and understandings can be inconsistent with teachers' intentions and can be impediments to further learning, it is important that teachers make attempts to understand how individual learners are thinking. Educational research (for example, McDermott, 1993) shows that even students who perform well on traditional forms of assessment (for example, applying a formula to solve a physics examination problem) can hold significant misconceptions (for example, believing that if an object is moving, there must be a force acting on it). The implications for assessment are that information is required about the *depth* of students' understandings. For diagnostic and teaching purposes, this involves gathering information about students' mental models and possible misconceptions. Beyond assessing factual and procedural knowledge, assessments are required of learners' abilities to transfer and apply concepts, principles and ideas to a range of relevant contexts.

The progressive nature of learning

Other research is providing insights into the nature of *learning progress* within particular learning domains. This research is answering such questions as: What is it that develops as individuals become more expert in a domain? What new knowledge, skills and understandings do they typically develop? What are common sequences and pathways of development? What are prerequisites for further learning? How does new learning build on and extend existing learning within the domain?

Some research studies have addressed these questions by investigating the development of increasingly deep understandings of specific concepts or phenomena. For example, Dawson and Stein (2008) explored the development of Year 9 students' understandings of two aspects of physics learning: the concepts of energy and forces/gravity. Based on a qualitative analysis of students' responses, they identified varying levels of understanding of these concepts, from the lowest, 8a, to the highest, 10b, as shown in Figure 2.8.

Dawson and Stein (2008) refer to these varying levels of understanding as 'descriptions of the pathways through which important energy concepts develop' (p. 90) and 'the conceptual space through which individuals move in their idiosyncratic ways' (p. 94). For each conceptual strand, there is a clear progression in students' levels of understanding. At the lowest levels (8a and 8b), students conflate the concepts of energy and movement; at the middle levels (9a and 9b), there is a belief that energy is something that causes motion; and at the highest levels (10a and 10b), there is an understanding that kinetic and potential energy are alternating energy states.

> *The most promising research on the development of science conceptions not only identifies correct and incorrect or novice and expert conceptions, but shows how conceptions develop over time. What we learn about the pathways through which concepts typically develop provides useful knowledge that can directly inform curriculum development.*

(Dawson & Stein, 2008, p. 90)

In the Figure 2.8 example, six levels of conceptual understanding were identified for each of these two physics concepts. But for any specific concept, the number of levels could be as few as two or three, provided that those levels make a useful contribution to describing the development of deeper conceptual understanding. The descriptions in Figure 2.8 contribute to an understanding of the nature of progress within the broader learning domain of Physics.

Figure 2.8: Levels of conceptual understanding of energy and forces/gravity

ENERGY		FORCES / GRAVITY
Students now fully grasp the idea that energy is the ability to do work. This translates into a more complete understanding of the relation between potential and kinetic energy, which are now treated as alternating energy states.	10b	Force and energy are consistently differentiated.
Students may claim that energy occurs in several forms and may explain (rather than simply stating) the idea that energy cannot be created or destroyed. Students may explain that energy is the ability to do work and describe multiple examples of energy doing work.	10a	Gravity is now viewed as a force that is involved in explanations of kinetic and potential energy. Force and energy are more differentiated than at single abstraction, though confusion may occasionally persist.
Energy is now rarely spoken of as though it is equivalent to motion. In defining energy, students may emphasise this point by referring to forms or sources of energy in which motion is not observable (electrical).	9b	Gravity is still largely viewed as a force that increases the energy of an object by increasing its speed. Concepts of energy and force are often poorly differentiated.
Energy is clearly viewed as 'something' that is 'behind' motion – a cause of motion. This notion is applied inconsistently. Sometimes, energy is still represented as equivalent to motion, especially when describing the energy of stationary objects.	9a	Forces acting on an object change its energy. For example, the energy of a dropped object increases due to gravity (a force). This is different from the representational systems argument that gravity makes an object fall or makes it fall faster. Force and energy are often confused. Students may interpret the definition of energy as 'the ability to do work,' as 'the ability to exert a force.'
Energy is something that pushes, pulls, or holds an object. Energy can be strong or weak.	8b	May make a connection between energy and force or gravity without being able to explain the connection. The terms, gravity, and force (when they are employed) are used, as the word energy is, to explain observed changes in motion.
Energy is the same thing as motion. Energy can be fast or slow. Energy is something you need for recess, hard work, etc. Energy moves things. Energy is 'in' an object.	8a	Gravity, if mentioned, is something that pushes, pulls or holds – like an invisible hand. Force, if mentioned, involves pushing, holding, or pulling on an object.

(Dawson & Stein, 2008, pp. 96-98)

Levels of conceptual understanding of the kind shown in Figure 2.8 represent *qualitatively* different understandings or ways of thinking about a phenomenon or aspect of a learning domain. Considered together, a set of levels for a concept is sometimes referred to as an 'outcome space', a term first used by Marton (1986). Some levels in an outcome space represent naive conceptions and partially formed schemas that characterise particular stages of learning (Pellegrino, Chudowsky & Glaser, 2001). As learning progresses, qualitatively different ways of representing and organising knowledge develop, enabling learners to solve a wider range of problems and to generalise their knowledge to a wider range of relevant contexts.

Importantly, conceptual levels of the kind illustrated in Figure 2.8 should not be taken to imply that all learners progress sequentially through a set of hierarchically organised levels. Individuals develop along idiosyncratic learning paths. Some may never show evidence of a particular way of thinking. Others may appear to regress in their understanding before developing higher levels of conceptual understanding. But the fact that individuals may not follow identical learning paths does not invalidate generalised pictures of increasingly sophisticated understandings of

an aspect of learning. Research-based learning progressions (or 'progress variables') provide valuable frames of reference for thinking about student learning and for establishing where individuals are at any given time in their progress toward deeper understandings.

> *A progress variable focuses on progression or growth. Learning is conceptualized not simply as a matter of acquiring more knowledge and skills, but as progressing toward higher levels of competence as new knowledge is linked to existing knowledge, and deeper understandings are developed from, and take the place of, earlier understandings.*

(Pellegrino, Chudowsky & Glaser, 2001, p. 115)

A second example of a set of qualitatively described levels is shown in Figure 2.9. These levels, referred to as 'proficiency levels' because they represent increasing levels of reading proficiency, were developed as part of the OECD's Programme for International Student Assessment (PISA) and are described and illustrated by Thomson, De Bortoli, Nicholas, Hillman and Buckley (2011).

Figure 2.9: Reading proficiency levels based on the OECD's Programme for International Student Assessment (PISA)

Level	Reading task description
6	Tasks at this level typically require the reader to make multiple inferences, comparisons and contrasts that are both detailed and precise. They require demonstration of a full and detailed understanding of one or more texts and may involve integrating information from more than one text. Tasks may require the reader to deal with unfamiliar ideas, in the presence of prominent competing information, and to generate abstract categories for interpretations. *Reflect and evaluate* tasks may require the reader to hypothesise about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text. A salient condition for *access and retrieve* tasks at this level is precision of analysis and fine attention to detail that is inconspicuous in the texts.
5	Tasks at this level involve retrieving information, require the reader to locate and organise several pieces of deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialised knowledge. Both interpretative and reflective tasks require a full and detailed understanding of a text whose content or form is unfamiliar. For all aspects of reading, tasks at this level typically involve dealing with concepts that are contrary to expectations.
4	Tasks at this level involve retrieving information require the reader to locate and organise several pieces of embedded information. Some tasks at this level require interpreting the meaning of nuances of language in a section of text by taking into account the text as a whole. Other interpretative tasks require understanding and applying categories in an unfamiliar context. Reflective tasks at this level require readers to use formal or public knowledge to hypothesise about or critically evaluate a text. Readers must demonstrate an accurate understanding of long or complex texts which may be unfamiliar.
3	Tasks at this level require the reader to locate, and in some cases recognise the relationship between, several pieces of information that must meet multiple conditions. Interpretative tasks at this level require the reader to integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. They need to take into account many features in comparing, contrasting or categorising. Often the required information is not prominent or there is much competing information; or there are other obstacles in the text, such as ideas that are contrary to expectation or negatively worded. Reflective tasks at this level may require connections, comparisons, and explanations, or they may require the reader to evaluate a feature of the text. Some reflective tasks require readers to demonstrate a fine understanding of the text in relation to familiar, everyday knowledge. Other tasks do not require detailed text comprehension but require the reader to draw on less common knowledge.

Level	Reading task description
2	Some tasks at this level require the reader to locate one or more pieces of information, which may need to be inferred and may need to meet several conditions. Others require recognising the main idea in a text, understanding relationships, or construing meaning within a limited part of the text when the information is not prominent and the reader must make low level inferences. Tasks at this level may involve comparisons or contrasts based on a single feature in the text. Typical reflective tasks at this level require readers to make a comparison or several connections between the text and outside knowledge, by drawing on personal experience and attitudes.
1a	Tasks at this level require the reader to locate one or more independent pieces of explicitly stated information; to recognise the main theme or author's purpose in a text about a familiar topic, or to make a simple connection between information in the text and common, everyday knowledge. Typically the required information in the text is prominent and there is little, if any, competing information. The reader is explicitly directed to consider relevant factors in the task and in the text.
1b	Tasks at this level require the reader to locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type, such as a narrative or a simple list. The text typically provides support to the reader, such as repetition of information, pictures or familiar symbols. There is minimal competing information. In tasks requiring interpretation the reader may need to make simple connections between adjacent pieces of information.

(Thomson et al., 2011, p. 22)

Unlike the qualitatively developed levels in Figure 2.8, the levels in Figure 2.9 were constructed from a statistical analysis of students' performances on a set of assigned reading comprehension tasks. On the basis of the statistical analysis, tasks were ordered from those most often answered correctly (Level 1b) to those least often answered correctly (Level 6) by 15-year-olds. Descriptions were then developed of the kinds of reading tasks located at each level.

When considered together, the seven levels in Figure 2.9 provide an empirically based map of increasing reading proficiency. In reality, these seven proficiency levels are simply convenient subdivisions of a continuum of increasing reading ability. Reading progress does not occur as jumps between discrete levels or bands, but incrementally over time. Most learning is best conceptualised as progress along a continuum that is not limited to any particular age or year of school. However, for the purposes of describing and reporting students' levels of proficiency, it is usual to mark out a continuum with a convenient numerical scale and/or to subdivide it into a number of levels or bands. In PISA, the underlying continuum is marked out using a convenient (but somewhat arbitrary) numerical scale and is also divided into seven broad 'proficiency levels'.

Importantly, the reading task descriptions in each level of Figure 2.9, like the level descriptions in Figure 2.8, are *empirically derived*. In other words, this picture of increasing proficiency is based not simply on somebody's intentions for, or beliefs about, the development of reading ability, but on an analysis of actual student reading performances. As such, it is not merely an intended curriculum sequence, but a research-based picture of increasing proficiency, which can serve as a basis for establishing where learners are in their learning and for monitoring the progress of individuals and groups of learners over time.

Proficiency scales of this kind also can be used to compare students' levels of achievement with the levels they are expected to have achieved or the levels required for particular purposes. This process of deciding on an appropriate level for a particular purpose is known as 'standard setting'. For example, following the development of the reading proficiency scale in Figure 2.9, the OECD undertook a standard setting process, which led to the identification of Level 2 as the minimum standard of reading proficiency necessary for someone to function effectively in the workplace and in life beyond school. The percentage of 15-year-olds in each participating PISA country not achieving at least Level 2 is identified and included in international reports of this assessment program.

> *Level 2 defines the level of achievement on the PISA scale at which students begin to demonstrate the reading literacy competencies that will enable them to actively participate in life situations. Students performing below this baseline are considered to be at serious risk of not achieving at levels sufficient to allow them to adequately participate in the 21st century work force and contribute as productive citizens.*

(Thomson et al., 2011, p. 22)

The concept of *progress* underlying these examples is perhaps the most fundamental concept in teaching and learning (Khoo, 2008). The goal of most teaching and learning is to develop deeper levels of understanding, increased levels of knowledge, more advanced skills and greater levels of expertise. Different terms are used to describe learning progress, including 'growth', 'development' and 'improvement'. Both students and teachers benefit from deep understandings of the nature of progress within an area of learning.

Teachers require deep understandings of progress to establish where students are in their learning, so that they can design learning opportunities appropriate to current levels of progress, set targets for improvement and monitor progress over time. School and system leaders also require good understandings of progress. For example, to interpret fully what it means for a certain percentage of students in a system to be performing 'below PISA Level 2', system leaders require an appreciation of the limited reading skills of students at that level. The substantive interpretation of student achievement is facilitated when assessments are made and reported against empirically based maps of learning progress.

The role of emotions in learning

Still other research is revealing the important role that emotions play in successful learning. Neuroscience research shows that healthy brain development depends not only on supportive physical environments (including good nutrition, adequate sleep and physical exercise), but also on supportive social and emotional environments. Animal studies reveal the importance to normal brain development of maternal affection and even physical contact. People are more likely to remember and learn if intrinsically motivated and emotionally engaged. Emotional engagement stimulates the brain to learn and individuals are more likely to remember when their emotions are aroused, when they are highly motivated and highly attentive (Organisation for Economic Co-operation and Development, 2007). On the other hand, research also reveals how negative emotions block learning. When individuals experience fear or stress – for example, when they are confronted with aggressive or negative behaviour or the prospect of failure – learning and memory are adversely affected.

Parallel learning research in educational settings has revealed the important role that classroom cultures and practices have in learning. Learning is more likely to occur when students are highly engaged and feel safe and supported in their learning, when teachers promote positive attitudes towards learning and encourage students' beliefs in their own capacity to learn, and when students are provided with meaningful feedback and are encouraged to engage in self-monitoring. In 'learning cultures' of this kind, learning tends to be intrinsically motivated – for example by curiosity, or by the challenge of finding solutions to meaningful problems. In contrast, 'performance cultures' are more likely to promote competition for success, and learning is more likely to be extrinsically motivated (for example, by external rewards or the fear of failure).

In her research into self-theories Dweck (2000) showed that successful learning is influenced by a belief in one's own ability to learn. More specifically, successful school learners tend to see ability as 'incremental' rather than 'fixed'. They believe that 'smart' is something you can become through effort, not something you are by birth. Students who view ability as fixed and who believe that they are naturally 'dumb' often believe that effort will make no difference. On the other hand, students who believe that they are naturally smart often underestimate the need for effort. Dweck showed that learners who view ability as incremental tend to have confidence in their own ability to learn, attribute lack of success to lack of effort, and remain optimistic that hard work will bring success.

This body of research, too, has implications for the assessment and reporting of student learning. In particular, assessment processes are likely to have an inhibiting effect on learning if they create unacceptable levels of stress for learners. Assessment processes are likely to be most effective in promoting learning if they enable self-monitoring and if they build self-belief by providing learners with evidence of the progress they are making over time.

In summary

Research in a number of disciplines is providing insights into basic learning processes and the conditions that support successful learning, and so is contributing to an evolving understanding of human learning. Some key observations from learning research include the following:

- Learning is often best conceptualised as an ongoing, potentially lifelong, process. Learners of the same age typically are at very different points in their learning and are progressing at different rates, but almost all learners seem capable of further learning progress if motivated to learn and if provided with learning opportunities appropriate to their current levels of progress and learning needs.
- Particularly important in most areas of learning is the development of schemas and conceptual understandings that allow learners to organise, transfer and apply their knowledge to new contexts. Deep understandings of these kinds develop only over extended periods of time, as learners gain experience with multiple examples of phenomena and opportunities to apply their learning in a range of situations.
- Most learning occurs as progress towards higher levels of proficiency or expertise – usually in the form of more extensive knowledge, more sophisticated understandings and higher-level skills. Empirically based understandings of typical pathways of increasing proficiency have an important role to play in informing teaching and monitoring learning progress.
- Learning is more likely to occur when positive emotions are aroused, when learners are highly motivated and highly attentive. Successful learning is also more likely when learners have strong self-beliefs in their capacity to learn, when they feel safe and supported in their learning, when they are given feedback that encourages and guides learning, and when they are able to see the progress they are making over time.

An emphasis on broader life skills and attributes

A third general pressure for assessment reform has arisen from the observation that commonly used assessment methods are of limited usefulness for assessing some of the skills and attributes now being identified as necessary for life and work in the 21st century. Many of these skills – for example, problem-solving, critical thinking, communicating, collaborating and self-management – are not unique to the 21st century, but have become more important skills and attributes in modern workplaces. In 21st century enterprises, teamwork, innovation and information sharing are growing in importance as more routine work processes are increasingly performed by technology (Binkley, Erstad, Herman, Raizen, Ripley, Miller-Ricci & Rumble, 2012). Employees need to respond flexibly to complex problems, to communicate effectively, to manage information dynamically, to work and create solutions in teams, and to produce new knowledge (Griffin, McGaw & Care, 2012). At the same time, the emergence of new digital technologies is requiring entirely new skills and ways of communicating and sharing information. Learning outcomes of these kinds are now being identified in education policy documents as key objectives of schooling in the 21st century.

The US National Research Council's Committee on Defining Deeper Learning and 21st Century Skills (Pellegrino & Hilton, 2012) reviewed a range of international efforts to identify competencies important for life and work. Their review clustered these competencies into three broad domains: cognitive, intrapersonal and interpersonal.

As a way to organise the various terms for '21st century skills' and provide a starting point for further research as to their meaning and value, the committee identified three broad domains of competence:

- *The Cognitive Domain includes three clusters of competencies: cognitive processes and strategies; knowledge; and creativity. These clusters include competencies such as critical thinking, information literacy, reasoning and argumentation, and innovation.*
- *The Intrapersonal Domain includes three clusters of competencies: intellectual openness; work ethic and conscientiousness; and positive core self-evaluation. These clusters include competencies such as flexibility, initiative, appreciation for diversity, and metacognition (the ability to reflect on one's own learning and make adjustments accordingly).*
- *The Interpersonal Domain includes two clusters of competencies: teamwork and collaboration; and leadership. These clusters include competencies such as communication, collaboration, responsibility, and conflict resolution.*

(Pellegrino & Hilton, 2012, p. 16)

In related work, Binkley et al. (2012) analysed relevant learning frameworks across a number of countries and identified ten skills that they clustered into four groupings:

Ways of Thinking
1 *Creativity and innovation*
2 *Critical thinking, problem solving, decision-making*
3 *Learning to learn, metacognition*

Ways of Working
4 *Communication*
5 *Collaboration (teamwork)*

Tools for Working
6 *Information literacy*
7 *ICT literacy*

Living in the World
8 *Citizenship – local and global*
9 *Life and career*
10 *Personal and social responsibility – including cultural awareness and competence*

(Binkley et al., 2012, pp. 18–19)

In Australia, the Melbourne Declaration on Educational Goals for Young Australians (Ministerial Council on Education, Employment, Training and Youth Affairs, 2008), renewed earlier declarations of national goals, and reiterated that the aim of schools should be to develop a broad range of life skills and attributes in their students. In the preamble to the Goals in the Melbourne Declaration is the following statement.

Literacy and numeracy and knowledge of key disciplines remain the cornerstone of schooling for young Australians. Schooling should also support the development of skills in areas such as social interaction, cross-disciplinary thinking and the use of digital media, which are essential in all 21st century occupations. As well as knowledge and skills, a school's legacy to young people should include national values of democracy, equity and justice, and personal values and attributes such as honesty, resilience and respect for others.

(Ministerial Council on Education, Employment, Training and Youth Affairs, 2008, p. 5)

Goal 2 of the Melbourne Declaration is 'All young Australians become successful learners, confident and creative individuals, and active and informed citizens.' Some examples of the student characteristics to be sought in pursuit of Goal 2, are listed under the following three headings:

Successful learners

- develop their capacity to learn and play an active role in their own learning
- have the essential skills in literacy and numeracy
- are creative and productive users of technology
- are creative, innovative and resourceful
- are able to think deeply and logically, obtain and evaluate evidence, solve problems, plan activities independently, collaborate, work in teams and communicate ideas.

Confident and creative individuals

- have a sense of self-worth, self-awareness and personal identity
- have a sense of optimism about their lives and the future
- are enterprising, showing initiative and using their creative abilities
- develop personal values and attributes such as honesty, resilience, empathy and respect for others
- relate well to others and forming and maintaining healthy relationships.

Active and informed citizens

- act with moral and ethical integrity
- appreciate Australia's social, cultural, linguistic and religious diversity
- have an understanding of Australia's system of government, history and culture
- are able to relate to and communicate across cultures, especially the cultures and countries of Asia
- work for the common good, in particular, sustaining and improving natural and social environments
- are responsible global and local citizens.

(Adapted from Ministerial Council on Education, Employment, Training and Youth Affairs, 2008, pp. 8–9)

The skills and attributes being identified through efforts of these kinds as being important to life and work in the future are usually not well addressed by traditional educational assessment processes, particularly by paper and pen assessments. In many cases, there is considerable work to be done to clarify and define these competencies as meaningful learning domains. Questions remain about the extent to which some skills and attributes (for example, problem-solving, decision-making, creativity) have meanings outside the context of specific school subjects. Is it possible to develop and assess these skills and attributes as generalised competencies, or can they only be developed and assessed meaningfully in particular areas of application? Other questions remain about the focus of assessments. For example, is the focus in assessing teamwork on how well an individual works in and contributes to a team, or is the work of the entire team the focus of the assessment? In practice, these assessment dilemmas remain largely unresolved. Despite the challenges in assessing many of the competencies now being identified as important, it is clear that most traditional assessment methods are inadequate to this task.

> *Traditional assessment methods typically fail to measure the high-level skills, knowledge, attributes and characteristics of self-directed and collaborative learning that are increasingly important for our global economy and fast-changing world.*

(Griffin, McGaw & Care, 2012, pp. v–vi)

Valid assessments of this broader range of skills and attributes will often require observations of learners' performances in complex situations in which they work collaboratively to solve real (or realistic) problems. Tasks will require students to apply what they have learnt to unseen problems, including by analysing situations, thinking critically, creating new solutions, communicating with others and making effective uses of available technology. Work to develop new forms of assessment is now underway in a number of centres, including through the international Assessment and Teaching of 21st Century Skills project (Griffin, McGaw & Care, 2012).

Importantly, the assessment of attributes and skills such as creativity, problem-solving, critical thinking, decision-making, teamwork, flexibility, initiative and communication require fundamentally different approaches from assessments designed to judge and report how much of a body of taught content students have successfully learnt. The purpose of these new forms of assessment must be to establish the levels of learners' skills/attributes at the time of assessment and to monitor improvements in those levels over time.

In summary

Globally, governments and education systems have placed a high priority on the development of skills and attributes necessary for life and work in the 21st century. However, very few of these competencies can be adequately assessed using the traditional methods designed to establish how well students have learnt what they have been taught, or to report in terms of judgements of success or failure. Instead, assessments of these skills and attributes will more appropriately be based on methods designed to establish students' proficiency levels at the time of assessment, recognising that these levels develop incrementally over extended periods of time.

The transformational potential of new technologies

The fourth general pressure for assessment reform arises from advances in technology. These advances are introducing the possibility of powerful new tools and approaches to assessing some forms of student learning. In the future, new technologies are likely to have a transformational impact on the field of assessment, including by enabling:

- assessments to take place anywhere, anytime, in ways consistent with an understanding of learning as an ongoing, lifelong process
- the creation of enhanced assessment environments through multimedia, greater interactivity, simulations etc.
- increased personalisation of assessments by tailoring assessment activities to individuals' readiness and needs, and by providing greater flexibility (on-demand assessments, remote delivery, etc.)
- assessments that are built from, and incorporate, scientifically based knowledge about the nature of learning in specific learning domains
- more automated explorations of students' thinking and solution strategies
- more automated, in-depth investigations and diagnoses of individuals' misunderstandings and errors
- high-quality, timely feedback to learners and others in forms that guide further action.

Advances in technology also have far-reaching implications for the collection, storage, analysis, interpretation and reporting of assessment information.

Enhanced assessment environments

Emerging technologies have the potential to transform assessment practice through the use of more interactive and intelligent forms of evidence gathering, including through simulations and virtual environments. The use of technology to create enhanced learning environments in Science was reviewed by Quellmalz, Timms and Schneider (2009). They conclude that significant advances have been made in the use of simulations in Science classrooms. In simulated learning environments students are able to manipulate variables such as forces, angles, distances and time and observe the effects of these changes, often in scenarios that are difficult or impossible to create in normal classroom environments, and that may not even be possible in the real world. In these 'virtual laboratories' students conduct experiments on-screen, replicating procedures and recording and analysing their observations and measurements electronically. In still more advanced applications, Science students are able to control the progress of personal 'avatars' through three-dimensional simulated environments. In simulations of these kinds, students can

interact safely with a wide variety of chemicals, scientific equipment and situations to which they might not otherwise be exposed.

> *Students can perform virtual experiments dealing with aqueous chemistry, such as acid–base reactions and solubility; physics experiments involving force and motion, springs, and electrical circuits; and virtual dissections of frogs and other animals. Virtual labs are also being used in introductory level college science and engineering courses to prepare students for work in real world labs in fields such as thermodynamics, robotics, and biotechnology.*

(Quellmalz, Timms & Schneider, 2009, p. 4)

Technology-enhanced learning environments can also provide unique opportunities to collect evidence about students' understandings of important principles and ideas in an area of learning. They provide opportunities to track the processes that students follow in attempting to solve problems and so provide a basis for assessing inquiry and problem-solving skills. However, it is still relatively rare to find environments of this kind being used as the basis for *assessing* student learning. Many science simulation projects do not use simulated environments to assess the understandings and skills that they are designed to develop, and some use paper and pen pre- and post-tests to assess student learning that occurs through simulations. In their review of projects that used enhanced learning environments in Science, Quellmalz, Timms and Schneider (2009) concluded:

> *In very few instances were assessments of student learning actually embedded within the simulation or designed to take advantage of its technological capabilities … Relatively scarce are clearly articulated descriptions of the evidence gathered to support claims of student learning. In most instances, rich streams of data from interactive tasks are not tapped as evidence of learning.*

(Quellmalz, Timms & Schneider, 2009, p. 4)

Although there are outstanding examples of the use of new technologies to build sophisticated systems for assessing human learning – including the use of computer-aided design as part of assessments for architects (Kenney, 1997) and the use of simulated patients in assessments for medical practitioners (Clauser, Margolis, Clyman & Ross, 1997) – most technology-based assessments to date have not capitalised on the potential of technology to transform assessment practice. In fact, most current computer-based assessment in school education is little more than paper and pen testing on a screen.

Personalised assessments

In addition to the potential for more interactive and intelligent forms of evidence gathering, technology-based assessments introduce the possibility of tailoring assessments to the achievement levels of individual learners. A feature of traditional standardised tests is that all learners are administered exactly the same set of test questions. However, as previously observed, students of the same age can be at very different points in their learning, meaning that a test designed for students in the 'middle' of a class will usually be too easy for more advanced learners and too difficult for less advanced learners. For both these groups of students, a test of average difficulty will be of limited use for establishing exactly where individuals are in their learning.

In a paper and pen context, one solution to this problem is to provide test items with a range of difficulties (so that all students are given at least some items at an appropriate level of difficulty) or to provide open-ended tasks that allow less advanced learners to engage and experience some success, and more advanced learners to demonstrate their higher levels of proficiency.

In an electronic environment, assessments can more easily be tailored to the achievement levels of individual learners. For example, when a bank of test items is stored electronically and a statistical estimate is available of the difficulty of each item in the bank, individual students

can be administered items appropriate to their current levels of achievement. In computer adaptive testing of this kind, items are selected automatically for administration based on an individual's performance on all items administered up to that point in the test. After a student's attempt at each item, the student's level of achievement is re-estimated and the item with the most appropriate difficulty for that level is selected from the bank. In this way, less advanced learners are administered easier tests, and more advanced learners are administered more difficult tests. Each student's results are then statistically adjusted for the difficulties of the items administered, enabling all students' estimated achievement levels to be compared directly on a described proficiency scale. The potential advantages of such assessments for teachers are more accurate information about where students are in their learning and more appropriate assessment experiences for individuals (rather than having some students attempting tasks which are much too easy or much too difficult for them).

Electronic assessments have the added advantage that, unless they are to be completed under supervised conditions, they can be undertaken anywhere, anytime. This means that learners usually can be assessed where and when they feel ready to be assessed and not at some pre-specified time and venue. In this way, electronic assessments enable learners to take a degree of control over their own learning and assessment.

Intelligent diagnosis

Another advantage of electronic assessments is that they have the potential to provide more detailed diagnostic information than is available from most paper and pen assessments. In an electronic environment there is the possibility of automatic recognition of possible student errors and misunderstandings. For example, if a student gives the answer 3/7 when prompted to add the fractions 2/3 and 1/4, an automatic hypothesis can be generated about the incorrect process that the student has followed. This hypothesis might be tested by prompting the student to add the fractions 1/2 and 2/5. If the student gives the answer 3/7, then the hypothesis regarding the incorrect process being applied might be considered confirmed, and an automatic inference drawn about the student's approach and level of understanding. The student's misunderstanding might then be flagged for the teacher's attention and, in an intelligent tutoring system, might lead to the delivery of appropriate electronic tutoring concerning the addition of fractions.

Intelligent diagnoses of this kind depend on electronic assessment systems which have built into them expert knowledge about the kinds of errors and misunderstandings that students commonly encounter and display in their learning. Electronic evidence gathering to explore and diagnose student misunderstandings depends on knowledge about how learning occurs in the domain being assessed. Intelligent tutoring systems are capable of providing feedback to students on their errors and giving hints to assist students in their learning. The kinds and levels of hints given, ideally, are determined by a student's responses and by the difficulties of the tasks to which they are responding (Timms, 2010). The kinds and number of hints that a student requires could also provide useful information for diagnostic purposes.

> *Technology-based systems have been developed to support individualized instruction by extracting key features of learners' responses, analysing patterns of correct and incorrect reasoning, and providing rapid and informative feedback to both student and teacher.*
>
> (Pellegrino, Chudowsky & Glaser, 2001, p. 10)

Rapid feedback

A further advantage of electronic assessments is that they have the potential to provide high-quality, rapid feedback to educational practitioners and to learners themselves. A feature of most paper-based assessment methods is a delay between the completion of an assessment activity and the receipt of feedback by students. Classroom assessments often require teachers

to read and assess written work and even computer-scored paper and pen tests take time to process. The advantage of electronic assessments is that they are capable of providing intelligent feedback both to students and their teachers, immediately after a set of tasks has been completed.

In summary

Electronic assessments have considerable potential to provide:

- rich assessment environments using multimedia, simulations and interactivity
- more informative assessment evidence, by tailoring assessment tasks to the readiness and needs of individual learners
- more useful diagnostic information, through the intelligent exploration of student responses
- more rapid and informative feedback.

This potential remains largely unrealised in most current technology-based assessments. The reasons are that, to deliver on their potential, electronic assessment systems tend to be expensive and difficult to build and require deep understandings of how learning occurs within a domain. Without the investment of time, expertise and resources to build intelligent assessment systems of this kind, there is a risk that the majority of electronic assessments will remain little more than collections of paper and pen assessments delivered on screen.

Concluding comments

This section has reviewed four general pressures for assessment reform, and has identified some common themes. For the purposes of educational decision-making, assessment methods designed originally for judging and reporting how well students had learnt defined bodies of curriculum content are less appropriate than methods designed to establish where learners are in their long-term learning progress (for example, methods such as diagnostic assessments of mathematics learning difficulties, and national assessments of average student reading levels). In general, decision-makers are less interested in judging student success than in understanding where students are in their learning and in identifying what can be done to promote further learning progress. For this reason, there is growing interest among educational decision-makers in such issues as: rates of student growth across the years of school, changes in average achievement levels over time, and measures of student achievement 'gaps' (for example, between Indigenous and non-Indigenous students and between students from different socioeconomic backgrounds).

Learning research reveals significant variability in achievement levels among learners of the same age, but also supports the view that every learner is capable of successful progress if motivated and provided with appropriate feedback and learning opportunities. In other words, research supports a view of learning as an ongoing, potentially lifelong process. Within any given area of learning, and at any given time, every learner is at some identifiable point in their long-term learning, with the capacity for further progress. Moreover, research shows that learners are likely to learn more successfully if they have positive beliefs about their own ability to learn. And one of the best ways to build these beliefs is to assist learners to see the progress they are making over time.

Other research has highlighted the importance to learning of deep understandings of concepts, principles and the big ideas of a learning area – the schemas that allow learners to organise their knowledge and to transfer and apply it to new contexts. Deep learning of this kind usually occurs only over extended periods of time, meaning that efforts to assess and monitor the development of understanding usually require long-term perspectives on learning.

Similarly, calls for schools to develop a broader range of life skills and attributes such as creativity, problem-solving, critical thinking, teamwork, flexibility, initiative and communication require long-term perspectives on learning and assessment. Assessments developed to judge student success in learning a body of curriculum content are manifestly inadequate for assessing skills and attributes of this kind. More appropriate are assessments to establish learners' skill

levels at the time of assessment and to monitor improvements in those levels over time.

Finally, as advances in technology introduce greater flexibility in where and how learning takes place, assessment methods designed for traditional modes of educational delivery will become increasingly irrelevant. More appropriate will be personalised assessments to establish where individuals are in their learning, to guide further learning, and to monitor individual progress over time.

These pressures for assessment reform require a more far-reaching response than the tweaking of traditional methods and approaches. They point to a need to reconceptualise educational assessment as the process of establishing where learners are in their long-term learning at the time of assessment. Under this conceptualisation, successful learning is defined as excellent learning progress (although there will often also be value in knowing how learners are performing in relation to other learners and/or age or year-level expectations). Assessments for this essential purpose require a coherent set of processes – an assessment 'system' designed specifically for this purpose. Important progress has been made in describing the elements of such a system, including by Pellegrino, Chudowsky and Glaser (2001) and Wilson (2005, 2009). Section 3 of this review builds on their work by outlining a set of design principles for a Learning Assessment System.

section

3 Design principles for a Learning Assessment System

Section 2 considered a number of current pressures to reform the field of educational assessment. These pressures are the result of growing desires to use assessment information to guide educational decision-making; evolving understandings of the nature of human learning; calls for schools to develop a broader range of life skills and attributes; and changes in where and how learning is taking place, particularly as a result of new technologies. These developments are demanding more than superficial changes to traditional assessment methodologies – they require a shift in how assessment is conceptualised and undertaken in school education.

Most traditional educational assessment takes place within a well-established paradigm. The starting point in this paradigm is the specification of what students are to learn. This is expressed in the form of a curriculum or course syllabus that spells out in varying degrees of detail the body of content – primarily, but not only, knowledge, skills and understandings – that teachers are to teach and students are to learn. These curricula and syllabuses are often developed by government agencies established for this purpose. With the curriculum specified, the role of teachers under this traditional paradigm is to teach the curriculum, the role of students is to learn what they are taught, and the role of assessment is to judge how much of what they have been taught students have successfully learnt. When undertaken during a course of instruction, 'formative' assessments may be used to judge how well students are learning, and 'diagnostic' assessments may be used to identify specific gaps in student learning.

Most efforts to reform educational assessment continue to operate within this basic paradigm. For example, they may encourage the use of a broader range of 'authentic' assessment methods, promote the use of assessment to inform teaching, or abolish external examinations in favour of school-based assessments, all with the same general intention of judging how well students have learnt (or are learning) what they have been taught.

As noted in Section 2, this traditional paradigm is not well placed to respond to current pressures for reform. The shift in approach outlined in this review is from *judging* student success in learning a body of specified content, to *understanding* – at various levels of detail – where students are in their long-term learning progress. In this sense, the use of assessment parallels the use of assessment in other professions where the focus is not so much on judging as on understanding.

This section describes a Learning Assessment System based on a set of five interdependent and mutually supporting design principles. Underpinning this system is a particular understanding of the purpose and nature of educational assessment; namely, that *the assessment process in education involves gathering evidence that can be used to draw conclusions about where students are in their long-term progress within a learning domain.*

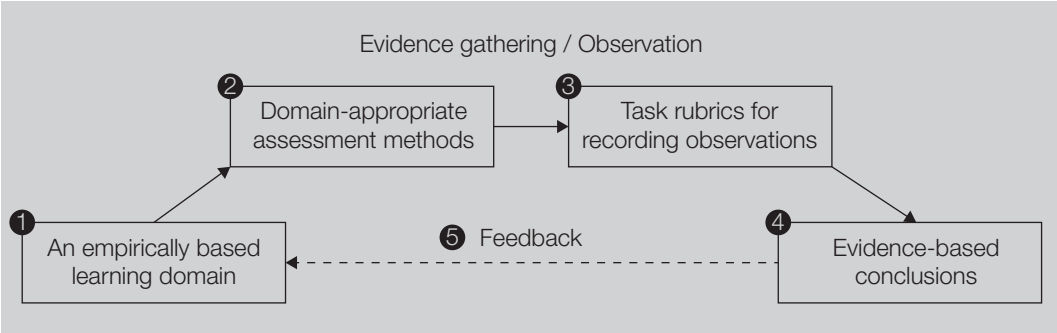
Several aspects of this understanding are worth noting. First, assessment always takes place in relation to an identified area (domain) of learning. Second, the essential purpose of this process is to draw a conclusion about where learners – either individually or as groups – are in their long-term learning progress within that domain. Third, there are two essential steps in this process: gathering evidence and then using that evidence to draw a conclusion. Pellegrino Chudowsky and Glaser (2001) refer to these two steps as 'observation' and 'interpretation' and they note that:

> *For an assessment to be effective, the three elements [a learning domain; an evidence gathering/observation process; and a process of drawing conclusions from evidence] must be in synchrony.*

(Pellegrino, Chudowsky & Glaser, 2001, p. 44)

As noted by Wilson (2005), the evidence-gathering process itself usually involves two sequential processes: first, deciding on an appropriate assessment method, and second, making records of observations and/or judgements of students' performances on specific assessment tasks. For example, to monitor writing progress, the assessment of samples of student writing is a domain-appropriate assessment method. Writing assessments are usually based on responses to specific writing tasks. But there is also the second-level question of what assessors will look for and record in assessing student writing. This question is addressed through the development and use of rubrics for judging and recording evidence (in this case) of student writing. These steps in the assessment process are illustrated in Figure 3.1. The broken line in this figure indicates that the final step in the assessment process is to provide feedback, including feedback that may contribute to an understanding of how learning occurs within the domain.

Figure 3.1: Five key steps in a Learning Assessment System



These five interdependent steps are referred to here as a Learning Assessment System. Each step in the system is underpinned by a design principle. These five steps and their underpinning principles are described, analysed and illustrated in the text that follows.

Empirically based learning domains

Step 1 in a Learning Assessment System is the specification and description of the learning domain to be assessed. This is the first step in any assessment process. The difference in a Learning Assessment System – and the basis of the first design principle – is that *the specification and description of the domain must be firmly grounded in research into the nature of learning within the domain*.

Most curriculum and syllabus development activities are located somewhere on a continuum representing varying degrees of research input. At one extreme are curricula that simply specify in a top-down way a body of content that students are to learn, with little or no input from research into how students learn that content. At the other extreme, but rare in practice, are curricula that are strongly grounded in domain-specific learning research. The approach being described here is towards the latter end of this continuum. To establish where learners are in their learning within a domain, it is necessary to have an excellent scientifically based

understanding of how learning typically occurs within that domain (an empirically based 'map' of the learning terrain through which learners progress).

The term 'domain' refers here to any area or aspect of learning to be assessed. Schools work to promote many different kinds of student learning. As well as developing knowledge, skills and understandings in particular subject areas, teachers work to develop a wide range of cross-curricular skills, attitudes and values. However, when it comes to assessing and monitoring the progress that students are making, it is usual to focus attention on just *one* area of learning at a time. This area may be relatively broad, such as Biology, or relatively narrow, such as human genetics. Attributes such as flexibility and creativity, and skills such as teamwork and collaborative problem-solving, are also assessed one at a time and are conceptualised here as 'domains' of learning.

Vertical structure of a domain

Especially important to an empirically based learning domain is an understanding and description of the nature of learning *progress* within the domain. What is the nature of increasing proficiency/expertise within the domain? What develops as learners become more expert? What are typical paths of development? How does new learning depend on, and build on, existing knowledge, skills and understandings? Are there prerequisites for some forms of progress within the domain? Are there common obstacles to further progress (for example, common errors and misunderstandings)? Questions such as these are answered through research into the nature of learning within the domain. This understanding of learning progress is referred to here as an understanding of the vertical structure of a domain.

These descriptions of learning progress are generally not tied to any particular age or year of school. Instead, they describe how learning occurs in the domain *largely independently of age or year level*, and usually over multiple years of learning. For example, the descriptions of increasing understanding and proficiency in Figures 2.8 and 2.9, although developed from an analysis of the responses of Year 9 students and 15-year-old students respectively, are likely to be equally useful for assessing and describing levels of understanding and proficiency of students in Years 10, 11 and 12. In contrast, traditional approaches to defining learning domains usually specify bodies of content to be learnt by students in particular courses or specified years of school.

Horizontal structure of a domain

As well as having a vertical structure, learning domains usually have a horizontal structure. The horizontal structure of a domain consists of the various aspects of learning that make up the domain. For example, each of the two aspects of physics learning in Figure 2.8 – energy and forces/gravity – could itself be conceptualised as a (relatively narrow) domain of learning, probably without a useful substructure. Alternatively, these two areas could be conceptualised as part of the substructure of a broader domain, perhaps 'conceptual understanding in physics', and so part of the horizontal structure of that larger domain. Similarly, PISA identifies a horizontal structure for the reading literacy domain, in part by separately assessing and reporting three Reading sub-domains: accessing/retrieving, integrating/interpreting and reflecting/evaluating.

Different approaches can be taken to defining the horizontal structure of a domain. One approach is to identify content subdivisions. For example, the domain of Mathematics is commonly conceptualised as comprising sub-domains such as number, measurement, space, probability and algebra. A different approach is to define the horizontal structure of a domain in terms of skill types and ways of thinking and working. For example, Kilpatrick, Swafford and Findell (2001) identify five 'strands' of mathematics learning, which they argue should be developed across all years of school.

- *conceptual understanding (comprehension of mathematical concepts, operations, and relations)*
- *procedural fluency (skill in carrying out procedures flexibly, accurately, efficiently, and appropriately)*
- *strategic competence (ability to formulate, represent, and solve mathematical problems)*
- *adaptive reasoning (capacity for logical thought, reflection, explanation and justification)*
- *productive disposition (habitual inclination to see Mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy).*

(Kilpatrick, Swafford & Findell, 2001, p. 116)

Ideally, the horizontal structure of the Mathematics learning domain might be conceptualised in terms of both mathematical content and mathematical skills simultaneously.

Scientifically based understandings of the vertical and horizontal structures of learning domains have important implications for curriculum design, teaching, learning and assessment.

Implications for curriculum design

In practice, much curriculum development is not based on solid empirical evidence about typical sequences and progressions of learning, but is driven more by conventions and beliefs about what should be taught in particular years of school. Wilson (2009) describes these latter approaches as a 'weak model' of curriculum development based on 'little more than a catalogue of desirable outcomes'. Too often, the result is a 'mile-wide-inch-deep' curriculum overcrowded with relatively superficial factual and procedural content (Schmidt, McKnight & Raizen, 1997). Much syllabus development is based only on horizontal considerations, resulting in lists of course-specific outcomes to be taught and developed in time-limited courses.

When curriculum design is informed by research into the vertical structure of a learning domain, teaching, learning and assessment take on a *developmental* perspective. The curriculum provides a frame of reference for planning and monitoring the development of more sophisticated understandings, higher levels of knowledge, more advanced thinking, improved communication skills, increased abilities to analyse and interpret information, and so on. The vertical structure of a domain makes explicit the nature of progress (or improvement) within that domain, usually over extended periods of time. Classroom teaching then takes place in the context of these longer-term understandings of learning. As Figures 2.2 to 2.4 make clear, because students at any given age have such varying levels of achievement, teachers require a good understanding of the nature of reading and mathematics development across multiple years of school.

> *The curriculum must be designed in terms of a model, grounded in evidence, of the paths through which learning typically proceeds … That is to say, the curriculum reflects and provides a strong model of progression in learning. This learning progression may then be used as a basis for both instruction and assessments.*

(Wilson, 2009, p. 7)

A focus on the vertical structure of a learning domain (that is, on learning across the years of school) is particularly important for the promotion of life skills and deep understandings that develop only over extended periods of time. Long-term learning within a domain often involves the development of more sophisticated understandings of subject matter, increasingly deep knowledge (for example, a growing appreciation of the contexts to which knowledge can be transferred and applied), and a developing ability to apply understandings and knowledge to real-world contexts. Failure to attend to the vertical structure of a domain in curriculum design can result in an under-emphasis on core skills and understandings of these kinds, and an overemphasis on isolated facts and procedures that do not build on prior learning, but instead constitute a 'catalogue' of desired outcomes.

> *An orientation toward core ideas focuses instruction and assessment on a comparatively small set of foundational concepts and traces a prospective developmental corridor – a pathway for learning across school grades and ages. A developmental corridor suggests that central concepts are introduced early in schooling and are progressively refined, elaborated, and extended throughout schooling.*

(Catley, Reiser & Lehrer, 2005, pp. 3–4)

Implications for teaching and learning

An empirically based learning domain provides a frame of reference for planning curriculum delivery, establishing where learners are in their learning and differentiating teaching to address the learning needs of individual learners. An understanding of typical learning paths within a domain is essential to monitoring learning progress and providing learning opportunities appropriate to students' current levels of achievement.

A shared understanding of the structure of a learning domain also facilitates conversations between teachers, students and parents about where individuals are in their learning, what progress they are making, and what actions might be taken to promote further learning. For these purposes, it is useful to consider both the range and kinds of skills, knowledge, understandings and attributes that make up the domain (that is, its horizontal structure), and also what it means to make progress towards higher levels of these competencies (that is, the vertical structure of the domain).

Research suggests that successful learning is more likely to occur when learners understand learning intentions. Rather than focusing only on the satisfactory completion of classroom activities, highly effective teachers explain and discuss the purposes of learning activities with students (Hattie, 2009). As a result, learners are able to explain what they are learning (for example, 'We are learning about the role of stereotypes in racism'). However, in practice, learning intentions are often communicated to students only as isolated objectives, divorced from larger understandings of learning progress.

Learning intentions are probably most effective when communicated in ways that allow learners to see what it means to make long-term progress and to set personal goals for improvement. When students are unclear about what it is that teachers are looking for (for example, in a piece of writing or a musical performance) they have no frame of reference for focusing their efforts. In extreme cases, the criteria against which their work or performance is assessed may remain mysterious, with students being left unsure about what they need to do to improve. Teachers can assist students in their learning by sharing with them examples of work (for example, samples of student writing or art work) at varying levels of quality and using these examples as a basis for discussing the characteristics of high-quality work and long-term progress. In this way learners are able to reflect on their current levels of achievement, set goals for improvement and monitor their own progress over time.

> *Students need to develop the capacity to monitor the quality of their own work during its actual production. For this to occur, students need to appreciate what constitutes work of higher quality; to compare the quality of their emerging work with the higher quality; and to draw on a store of tactics to modify their work as necessary … Many students whose usual levels of performance are mediocre are hampered by not knowing what constitutes work of high quality. This sets an upper bound on their ability to monitor the quality of their own developing work.*

(Sadler, 2009, pp. 1–4)

Implications for assessment

An empirically based learning domain also provides a frame of reference for assessing student learning. Assessments of student learning within a domain must first provide adequate coverage of the horizontal structure of the domain. For example, assessments of mathematics learning

usually would be considered to lack 'construct validity' (Messick, 1994) if they did not adequately sample learning from various sub-areas of Mathematics and/or major mathematical skills such as those identified by Kilpatrick, Swafford and Findell (2001).

In practice, educational assessments often address *only* the horizontal structure of learning domains. These assessments are designed to sample the domain substructure to enable conclusions to be drawn about how much of the body of taught content (facts, concepts, skills, procedures, etc.) students have successfully learnt. These conclusions are generally reported as percentages or grades.

In contrast, assessments designed to establish where learners are in their ongoing learning progress explicitly address *both* the horizontal and vertical dimensions of a learning domain. Horizontal coverage of the domain remains important, but assessments are now made not simply against a body of taught content, but against an empirically based understanding of learning progress within the domain. Assessments of this kind provide information about the points individuals have reached in their learning at the time of assessment and the progress they make over time. Masters and Forster (1997b) refer to such assessments as 'developmental' in nature. Assessments of this kind can be undertaken at the beginning of, during, or at the end of a course of instruction, or without reference to a course of instruction at all. Because the essential purpose is the same at these different times, there is no reason why the assessments themselves should be of fundamentally different kinds. Any given assessment can be used both to identify starting points for future teaching and learning (that is, function as an assessment *for* learning) and to provide information about learning progress since some earlier occasion (that is, function as an assessment *of* learning).

> *Most assessments provide 'snapshots' of achievement at particular points in time, but they do not capture the progression of students' conceptual understanding over time, which is at the heart of learning. This limitation exists largely because most current modes of assessment lack an underlying theoretical framework of how student understanding in a content domain develops.*

(Pellegrino, Chudowsky & Glaser, 2001, pp. 27–28)

In summary

An empirically based learning domain is distinguished by the fact that it:

- is strongly grounded in learning research, rather than being based simply on specified teaching and learning intentions
- describes learning progress within the domain, rather than merely listing intended learning outcomes
- is not time-limited to a particular course or year, but describes the nature of long-term (multi-year) learning.

As such, an empirically based learning domain provides a map for establishing where students are in their long-term learning and monitoring progress over time. When teachers, parents and students share an understanding of progress within a domain, learning goals and intentions can be set and discussed in the context of this larger understanding. And the assessment of student learning moves from being a process of judging student success to a process of understanding where students are in their learning at any point in time.

Domain-appropriate assessment methods

Step 2 in a Learning Assessment System (see Figure 3.1) is the choice of a general assessment method and the design of specific assessments tasks for gathering evidence about the domain of interest. The design principle underpinning this step is that *the assessment method must be designed to provide useful information about where learners are in their learning within the domain*. There are several relevant design considerations.

Construct validity

A first consideration in developing domain-appropriate assessment tasks is the choice of an appropriate general method of assessment. Different assessment methods are valid for different kinds of learning. For example, learning in areas such as dance, drama, instrumental music, oral language, oral reading and physical education can be validly assessed in part through direct observations of student *performances*. In other areas of learning, valid assessments require observations of things that students make – the *products* of student work, including works of art (paintings, drawings, photographs, sculptures, films, etc.) and works of technology (metal, ceramics, wood, food, textiles, etc.). Student work is sometimes brought together in a *portfolio* of evidence. When assembled over a period of time, portfolios can provide a valid basis for establishing current levels of achievement and for monitoring progress over time. The ability to plan investigations, gather, analyse and evaluate relevant information, synthesise findings, and communicate conclusions is often validly assessed through extended student *projects*. And many kinds of learning can be validly assessed using written tasks, tests, exercises, essays and assignments, completed either in paper and pen form or electronically.

All these and other general assessment methods are capable of providing valid information about particular kinds of learning. The most important consideration in choosing an assessment method is the method's capacity to provide information about where students are in their learning within the domain of interest – in other words, its construct validity or fitness for purpose.

Another way of saying this is that no assessment method is *inherently* more valid than any other; validity can be evaluated only with reference to the learning domain and the assessment purpose (for example the level of diagnostic power required). It is not uncommon in the assessment literature to see particular assessment methods promoted as inherently preferable to other methods. For example, 'authentic' and 'performance' assessments may be described as more valid than standardised tests. This generalised preference is, in part, a response to the over-use of multiple-choice testing in the United States of America. In that country, 'performance' assessments often are interpreted as anything other than multiple-choice tests. Similarly, 'authentic' assessments involving real-life problems often are described as superior to assessments based on invented tasks. The position taken in this review paper is that different assessment methods are appropriate for different aspects and domains of learning. For example, complex, real-life mathematics problems may be particularly effective for the assessment of students' abilities to apply their mathematical understandings, but may be an inefficient way of investigating and diagnosing specific mathematical errors and misconceptions.

In educational settings, construct validity is often interpreted as adequate coverage of a specified body of curriculum content, or in the terminology introduced above, adequate coverage of the horizontal structure of a learning domain. But assessments of the kind described and advocated here also must provide information about where learners are in their long-term learning progress through a domain. To be construct valid, such assessments must explicitly address both the horizontal and vertical dimensions of a domain.

An assessment method must also be capable of supporting the next essential steps in the assessment process – that is, making systematic records of students' responses or performances and reaching conclusions about where students are in their learning within the domain of interest. If assessments are to be useful for these purposes, then they must be selected carefully. They cannot be based simply on personal preferences – for example, a preference for informal classroom observations over performances on externally developed tasks – but must be chosen and designed to provide evidence that permits meaningful conclusions.

> *The tasks to which students are asked to respond on an assessment are not arbitrary. They must be carefully designed to provide evidence that is linked to the cognitive model of learning and to support the kinds of inferences and decisions that will be based on the assessment results.*
>
> (Pellegrino, Chudowsky & Glaser, 2001, p. 47)

Having decided on a general assessment method, a decision is also required about the specific tasks to which students will respond. Tasks could be individual exercises such as problems to be solved, procedures to be performed and questions to be answered, or large and complex activities such as extended research projects. Tasks define the specific contexts in which assessment information is to be gathered.

Although assessment tasks must be carefully designed, individual tasks are rarely of intrinsic interest, but are best viewed as a means to an end. Conclusions drawn in Step 4 of a Learning Assessment System (see Figure 3.1) are *inferences* about where students are in their learning based on *interpretations* of the available evidence. For example, in an assessment of reading, assessment tasks are used to gather evidence about students' levels of reading proficiency. However, students may never again have to read and answer questions about the specific pieces of text used for this purpose. Those pieces of text and their associated questions are simply vehicles for the collection of relevant evidence about what is really of interest – a student's underlying level of reading proficiency. In educational assessments, individual tasks generally are transient and interchangeable and are important only to the extent that they permit meaningful inferences about levels of proficiency and progress within the domain of interest.

Reliability

A second consideration in developing domain-appropriate assessment tasks relates to the desired *precision* of conclusions about where learners are in their learning within a domain. In general, the desired level of precision will depend on the use(s) to which the results of the assessments are to be put. For some uses, such as measuring the progress a learner has made over time or measuring national trends in student achievement levels, relatively precise estimates may be required. For other uses, such as establishing general starting points for teaching, less precise estimates may suffice. Other terms for precision are 'confidence' and 'reliability'. The more precise the estimates of where learners are in their learning, the greater the confidence that can be placed in those conclusions or, alternatively, the greater their reliability.

The precision or reliability of assessment conclusions depends on the *amount* of domain-appropriate evidence upon which they are based. In general, the level of confidence that can be placed in assessment conclusions increases with the amount of evidence on which those conclusions are based. For this reason, educational assessments are almost always based on multiple pieces of evidence. Knowing how students performed on a single task, unless it was a particularly large and complex task, is usually an inadequate basis for establishing where students are in their learning within a domain. Similarly, a handful of test questions is not usually an adequate basis. As the number of domain-relevant assessment tasks is increased, so is the level of confidence in the assessment conclusion.

However, the precision of assessment conclusions also depends on the *appropriateness* of the assessment tasks for the individuals being assessed. For example, if a student is given a set of tasks which are all much too easy (or much too difficult) for them, then those tasks will not be particularly helpful in pinpointing where the student is in his or her learning. Although the tasks may be appropriate for other students, they provide little information about, and thus provide an imprecise estimate of, where that particular student is in his or her learning. The uncertainty about a learner's exact location within a learning domain is usually indicated in terms of a region of uncertainty (or 'confidence interval') around the best estimate. The higher the level of confidence in the estimate, the narrower the region of uncertainty about that estimate. Confidence intervals also can be reported around group means.

Objectivity

A third consideration in developing domain-appropriate assessment tasks is the 'objectivity' of the resulting assessment conclusions. The conclusions reached from an assessment process should be 'objective' in the sense that they do not depend on which specific assessment tasks are used or who does the assessing. Objectivity is a familiar idea in everyday assessments. For example, we expect

measures of blood pressure or temperature not to depend on which particular instrument is used to obtain them or who does the measuring. Similarly, when our interest is in assessing reading comprehension, we intend conclusions about levels of reading comprehension not to depend on which particular reading passages are used, which questions are asked about those passages, or who does the assessing. In other words, we expect the outcomes of the assessment process to be 'objective'.

Objectivity is particularly important when comparisons are to be made across different sets of assessment tasks. To monitor the progress students make over time, it is necessary to compare assessments made on different occasions. In classrooms, these might be assessments made at the beginning of the school year and again towards the end of the year, usually using different assessment tasks. In national and international achievement surveys, assessments are made on a multi-year cycle, with different assessment tasks being used in each cycle. In both cases, the intention is to compare results on different sets of assessment tasks. The difficulty, however, is that it is not possible to compare directly raw scores on different sets of tasks. For example, a score of 20 on a test of 30 relatively easy questions does not have the same meaning as a score of 20 on a test of 30 more difficult questions. Raw scores are not 'objective' because they do not have meanings independent of the instruments used to obtain them.

One approach to achieving objectivity in educational assessments is to make statistical adjustments to students' results for the difficulties of the tasks they attempt. The goal of this process is to estimate where students are on the same described proficiency scale, even if they attempted different sets of assessment tasks. For example, a score of 510 on the PISA reading scale is intended to represent the same level of reading proficiency in each PISA cycle, regardless of the difficulties of the reading tasks used in any particular cycle. This statistical approach also is used by national and international assessment programs such as NAPLAN, TIMSS, PIRLS (the Progress in International Reading Literacy Study) and the ICCS (the International Civic and Citizenship Education Study; Schulz, Ainley, Fraillon, Kerr & Losito, 2010) and by commercial assessment resources such as the Progressive Achievement Tests (PAT) (ACER, 2008) and the TORCH Tests of Reading Comprehension (ACER, 2003). In all these assessment processes, students' performances on, or responses to, assessment tasks are analysed statistically and used to infer students' locations on a described proficiency scale.

A second approach to achieving objectivity is to work to achieve high levels of consistency (sometimes called 'inter-rater reliability') in assessors' interpretations and uses of described proficiency scales. Under this approach, assessors make global, impressionistic judgements of where students are on a scale. The process is a demanding one, requiring assessors to mentally weigh the available evidence and make an on-balance judgement. Judgements of this kind are 'objective' to the extent that they do not depend on the specifics of the assessment contexts or who does the assessing. Assessments will not be objective if they are influenced by differences in assessor harshness/leniency or by factors such as an assessor's prior knowledge of a student.

Teachers can use both these approaches in classroom assessments. First, by using professionally developed assessment resources that adjust for the difficulties of the tasks that students undertake, objective assessments can be made for some aspects of student learning. For example, PAT Reading tests increase in difficulty across the years of school. However, because the results of all PAT Reading tests are expressed on the same reading proficiency scale, teachers can use different tests to monitor students' reading progress/growth over time. Second, by collaborating with other teachers – for example, through moderation activities – teachers can increase the consistency and comparability of the assessments they make against described proficiency scales. For example, when teachers judge student progress against the levels of a developmental continuum in Writing, and when their judgements are consistent across different samples and kinds of student writing, and are consistent with the judgements of their colleagues, then writing assessments can be said to be made 'objectively'.

Inclusivity

A fourth consideration in developing domain-appropriate assessment tasks is the extent to which the assessment process provides useful information about the achievements and progress of *all*

students with whom it is to be used. An assessment process should not provide underestimates or overestimates of some students' levels of achievement or progress because of their gender, physical disability, cultural background or geographical location.

Teachers require an understanding of how the contexts and language of assessment activities can provide underestimates or overestimates of some students' levels of achievement. The topics on which students are asked to write and the contexts in which tasks are set can place some students at a disadvantage. For example, girls often perform less well on activities related to sports than they do on other kinds of activities. Assessments that assume experiences that some students may not have had – for example, metropolitan train travel or familiarity with oceans and beaches – can also lead to underestimates of student achievement. Care is required to ensure that classroom assessment activities do not disadvantage students from particular cultural backgrounds (for example, by expecting eye contact during oral language assessments), are not offensive to particular religious groups, and do not disadvantage some students by evoking emotional reactions (for example, by avoiding references to terrorism, death and violence).

Professional test development processes include checks for the possibility of test question 'bias'. Sensitivity reviews are undertaken to inspect questions for features that may place particular groups of students at a disadvantage. Doubtful questions are removed before the test is finalised. Further checks can be conducted after a test has been administered. If a statistical analysis of students' test performances identifies some questions as being unexpectedly difficult (or easy) for particular subgroups of students, and if a substantive reconsideration of those questions raises doubts about their fairness, then responses to those questions can be removed prior to reporting assessment results.

Feasibility

A final consideration in developing domain-appropriate assessment tasks is feasibility or practicability. Although a particular assessment method may be ideal from the point of view of its construct validity, it may not be particularly feasible in practice. For example, in some areas of learning, the most valid conclusions about individual learning might be based on extended, in-depth interviews. However, such interactions may not always be feasible in practice. In all assessment contexts, practicability is an important consideration that needs to be addressed alongside considerations of validity, reliability, objectivity and inclusivity.

In summary

Step 2 in a Learning Assessment System is the choice of a general assessment method and the design of specific assessment tasks to collect evidence about where learners are in their learning progress, within the domain of interest. Crucially, the general assessment method must be construct valid. In other words, it must be strongly aligned with an empirically based understanding of the learning domain. In practice, the five considerations described above – validity, reliability, objectivity, inclusivity and feasibility – usually cannot be maximised simultaneously, so compromises and trade-offs must be made. Some considerations are more important in some assessment contexts than in others.

Rubrics for recording observations

Step 3 in a Learning Assessment System (see Figure 3.1) is the development of task rubrics for recording students' responses to, or performances on, assessment tasks. Records of observations are important because they provide the evidence for the next, interpretive stage in which overall conclusions are drawn about where students are in their learning. The principle underpinning this step is that *students' responses or performances should be recorded using one or more sets of categories (rubrics) which are task specific, hierarchical, and qualitatively defined*.

Records of observations are essential because, without them, conclusions depend too heavily on memory and are also less likely to be made consistently and reliably. Records also provide details that can be important for diagnostic purposes and for guiding future action. In some

assessment contexts, written notes are kept during the observation of student performances (for example, a drama production or gymnastics performance) or while inspecting student work. These notes typically record specific features of a performance or piece of work and serve as an aide mémoire. However, the records made in a Learning Assessment System are more than informal notes. They are based on a 'rubric' – a set of described and ordered categories for recording students' responses to, or performances on a task. These categories represent increasing quality or sophistication of response to a task.

A very simple rubric consists of only two levels. For example, when responses to a question are recorded as either right or wrong, the rubric for that question consists of two levels, with one level representing a higher level of response than the other. Other rubrics consist of more than two ordered levels. Multi-level rubrics are commonly used to record responses to complex assessment tasks. For example, rather than recording attempts at a problem as either right or wrong, partial credit is commonly assigned for partially correct responses or partial success, resulting in a rubric with three or more described levels. Similarly, student essays, projects, instrumental music performances and gymnastics routines typically are judged against rubrics consisting of several ordered levels of performance.

Importantly, the levels that make up a rubric are *qualitatively defined*. Even for the simplest of tasks scored right or wrong, prior decisions must be made about which responses will be treated as 'right' and which will be treated as 'wrong' (For example, will 2 + 2 = ㅓ be scored as right or wrong? Will 2 + 2 = ≣ be scored as right or wrong?). The outcome categories for all assessment tasks must be predefined qualitatively.

> *Inherent in the idea of categorisation is an understanding that the categories that define the outcome space are qualitatively distinct. In fact, all measures are based, at some point, on qualitative distinctions. Even fixed-response formats such as multiple-choice test items and Likert-style survey questions rely on a qualitative understanding of what constitutes different levels of response.*

(Wilson, 2005, p. 63)

As well as providing a basis for evaluating and recording students' responses to an assessment task, a rubric also communicates expectations for students' responses and work. A rubric makes transparent what assessors are looking for – the criteria that they use in evaluating students' responses to a task or the characteristics of high-quality work.

In evaluating and recording performances on large and complex assessment tasks, it is common to assess multiple aspects of performance. In these situations, a rubric is developed for each aspect separately. Assessments of this kind are sometimes referred to as 'analytical' assessments because they analyse complex performances into a number of aspects.

Task-specific rubrics

The rubrics used in a Learning Assessment System have several characteristics. First, they are task specific. Each rubric consists of a set of levels that describe increasing quality or sophistication of response to a particular task.

Each level of a rubric describes task-specific performances at that level. These levels may be labelled for convenience (for example, 1 / 2 / 3 / 4), but a rubric is more than a set of labels; it is a set of descriptions of observable features of performance. Some attempts to develop rubrics use more or less the same general description of performance at each level, qualified by ordered adjectives such as 'limited', 'sound' and 'excellent'. Rubrics in a Learning Assessment System do more than rely on the interpretation of ordered adjectives to describe increasing performance on a task; they describe *observable characteristics* of performance at each level.

Sometimes, the same general rubric can be used with different assessment tasks. For example, a rubric for the assessment of students' writing may be essentially the same regardless of the topic on which students are asked to write. However, even in these situations, general rubrics must be interpreted in the context of specific tasks. For example, the application of a general writing rubric to a particular writing task can often be usefully illustrated with samples of students' writing at each level of the rubric.

Even when categories are superficially the same from context to context, their use inevitably requires a re-interpretation in each new context.

(Wilson, 2005, p. 67)

Hierarchically organised categories

Considered together, the levels of a rubric form a hierarchy, and so help to describe in operational terms, the direction of progress within the learning domain. This is true even if a rubric consists of only two levels, with one level indicating a higher level of performance than the other. Because rubrics provide an operational definition of progress, they must be consistent with what is known from research about the vertical structure of the domain.

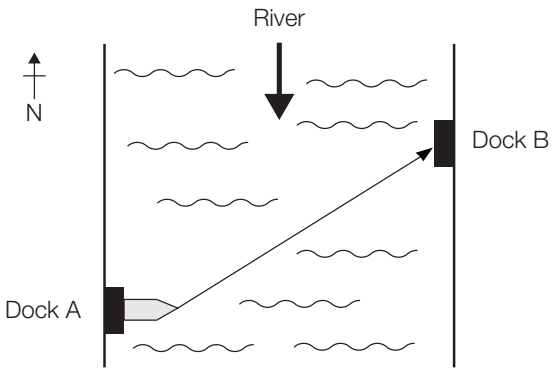
Section 2 of this review considered examples of learning progressions constructed from learning research. Those progressions described increasing conceptual understanding of energy, increasing conceptual understanding of forces/gravity (Figure 2.8) and increasing reading proficiency (Figure 2.9). None of these learning progressions was task specific. Each had been generalised from multiple observations of students' responses. However, once learning progressions of this kind have been constructed, specific assessment tasks can be developed to gather evidence about where students are in relation to a learning progression. The development of assessment tasks and their rubrics is ideally informed by what is already known about the nature of progress within the relevant learning domain.

In practice, rubrics are often developed not with reference to a pre-existing description of a learning progression, but on the basis of expert knowledge of a domain. For example, in assessing students' solutions of a problem, a teacher may judge some solution strategies to demonstrate more sophisticated understandings than others and award greater credit to these superior solutions. In this way, the teacher uses their expert knowledge to identify varying levels of response to the task, even though this expert knowledge may be more implicit than explicit.

An attempt to construct a task rubric that makes explicit students' varying levels of understanding is described by Ramsden, Masters, Stephanou, Walsh, Martin, Laurillard and Marton (1993). They gave physics students a task based on a motorboat crossing a flowing river (Figure 3.2). Students were asked to provide written responses to this task, which were then analysed to identify students' varying explanations and understandings.

The analysis of students' responses to this task resulted in the identification of five qualitatively different kinds of response, which were then ordered from those reflecting the lowest level of understanding of the necessary physics (Figure 3.3, Category E) to those reflecting the highest (Category A). The resulting rubric for this task is itself the result of qualitative research and contributes to the description of increasingly sophisticated understandings of this aspect of physics learning.

Figure 3.2: Motorboat task



A motorboat with its engines running at a constant rate travels across a river from Dock A to Dock B in a straight line, as shown above. Compare the times taken for this journey when the river is flowing and when it is not. Fully explain your answer.

(Ramsden, Masters, Stephanou, Walsh, Martin, Laurillard & Marton, 1993, p. 304)

Figure 3.3: Rubric for motorboat task

Category	Description
A	Response is based on an understanding that longer distance and same speed imply longer time. The focus of the response is on *distance relative to the river*. Two frames of reference (ground and river) are explicitly distinguished.
B	Response is based on an understanding that lower speed relative to the ground implies longer time. The focus of the response is on *speed*. The student recognises that, in the condition where the river is flowing, the boat's velocity is smaller and the displacement is the same. There is an implicit acceptance of more than one frame of reference.
C	Response is based on an understanding that longer distance implies longer time. The focus of the response is on *distance covered*. Frame of reference (the ground only) and speed are taken for granted. The boat's path may be described as parabolic or discontinuous.
D	Response is based on an understanding that same distance implies same time. The focus of the response is on the *path travelled*. The frame of reference and speed are taken for granted.
E	Response is based on an understanding that less pushing force implies longer time. The focus of the response is on a *dynamic explanation only*. A linear relation between speed, distance and force is assumed; connections between force and velocity are taken for granted.

(Ramsden, Masters, Stephanou, Walsh, Martin, Laurillard & Marton, 1993, pp. 306–7)

Regardless of how it is constructed, a task rubric consists of a hierarchy of levels, which relate to the vertical structure of a learning domain. These levels may be informed by existing descriptions of learning progress within the domain or by teachers'/assessors' implicit understandings. A well-developed task rubric can itself contribute to the description of progress within the domain.

Qualitatively defined categories

A task rubric is a construction in the sense that, in developing a rubric, decisions are required about how many levels of response to a task will be identified and how those levels will be described. There is no pre-existing or correct number of levels; the primary requirement is that the levels of a rubric usefully differentiate different levels of performance or response.

Distinctions between the levels of a rubric need to be described qualitatively in such a way that the characteristics of each level are clear to rubric users. Conversations about a rubric and its interpretation and use can also help in clarifying distinctions between levels. The process of formalised discussions of this kind between users is sometimes referred to as 'moderation'. The purpose of moderation meetings is to reach shared understandings of task rubrics with a view to ensuring consistency in their application. The process sometimes includes teachers bringing assessments that they have made of their own students' work for discussion with other teachers.

Also helpful in this moderation process are examples of students' responses or work that illustrate performance at each level of a rubric. These 'exemplars' are concrete examples that assist in clarifying the meaning of verbally described levels. Some excerpts from students' written responses to the powerboat task are shown in Figure 3.4. A physics teacher wishing to use this task may be assisted in their interpretation and application of the rubric by these (and other) illustrations of student thinking at each level.

Figure 3.4: Examples of student explanations (powerboat task)

Category	Description
A	'If the river is flowing, it is going to have to overcome the river's velocity before it makes any headway upstream… It has to travel further north relative to the river but not relative to the docks. Relative to the water, it will have travelled further.'
B	'When the river is flowing, the boat needs to head for a spot further up the river than the dock. The velocity of the boat in this initial direction will be the same as when the river was not flowing, but because it will be swept down the river, its actual velocity will be less: this will cause the time to be much longer.'
C	'You would probably have to have the captain steering the boat up at that angle in order to get the situation where he goes upriver, but the flow of the river takes him back downriver. [Draws line with sharp angles alternating right and left.] So you go up and then come back down. So you travel a further distance to do the same thing.'
D	'You've got a constant velocity and a constant distance. Neither of these are changing. The time is going to be exactly the same.'
E	'When the river is flowing, the time will be greater because when the river's opposing it, the boat has to angle upstream, so all the thrust is not used up getting across. Some of it's used up trying to oppose the force of the river.'

(Ramsden, Masters, Stephanou, Walsh, Martin, Laurillard & Marton, 1993, pp. 309–10)

In summary

Step 3 in a Learning Assessment System is the development of task rubrics for recording responses to assessment tasks. Rubrics are at the heart of the assessment process because they make explicit what is being looked for and valued as evidence of successful learning. The levels of a rubric contribute to the operational definition of progress within the learning domain. They describe increasing quality or sophistication of response, are developed to differentiate levels of response, and are most useful when described qualitatively and illustrated with examples.

Evidence-based conclusions

Step 4 in a Learning Assessment System (see Figure 3.1) is the drawing of conclusions about where learners are in their progress within a learning domain, on the basis of collected and recorded evidence. The design principle underpinning this step is that *conclusions should be drawn with reference to an explicit, empirically based understanding of learning progress within the domain*.

The process of drawing conclusions about where learners are in their learning depends on a deep understanding of the learning domain itself, including a knowledge of how learning typically progresses within the domain – for example, an understanding of prerequisite skills and knowledge for successful further learning and an awareness of common misunderstandings, errors and obstacles to learning progress. Deep knowledge of this kind involves more than familiarity with an intended curriculum. It depends on professional training, experience and accumulated research into the nature of learning within the domain.

For example, drawing conclusions about where students are in their early reading development requires a deep understanding of the domain of Reading. This includes an understanding of typical paths of reading development. A number of researchers have developed empirically based maps of reading development (for example, Chall's (1996) six stages of reading development). But it also includes a deep understanding of the skills of Reading (for example, phonemic awareness and phonological awareness) and of impediments to reading success (for example, limited oral language development, limited vocabulary, poor metacognitive skills). The process of establishing where individuals are in the reading development may include an overall assessment of their level or stage of reading development, but also may require more detailed analysis and diagnosis of specific pre-reading and early reading skills.

Maps of learning progress

An essential feature of any empirically based learning domain is its description of long-term learning progress within the domain – referred to earlier as the 'vertical' structure of the domain. Rather than being catalogues of desired learning outcomes, empirically based learning progressions describe what it means to make ongoing learning progress (for example, by describing the nature of increasing expertise, deeper understandings or higher-order skills).

Examples of empirically based progressions were provided in Figures 2.8 and 2.9. Further examples are provided in Figures 3.5 and 3.6. The spelling continuum in Figure 3.5 was developed by the Education Department of Western Australia to assist teachers to make judgements about where students are in their spelling development. Descriptions of spelling development were 'extracted from research into the development of literacy in English-speaking children' and clustered into five ordered phases of increasing spelling proficiency. Teachers make judgements based on their observations of students' spelling behaviours and with reference to the 'indicators' and accompanying illustrations of each phase of the continuum (Education Department of Western Australia, 1994).

Figure 3.5: First Steps Spelling Developmental Continuum

Phase	Description
5 Independent Spelling	In this phase writers have become aware of the many patterns and rules that are characteristic of the English spelling system. When spelling a new word they use a multi-strategy approach. They have the ability to recognise when a word doesn't look right and to think of alternative spellings. Spellers in this phase will have accumulated a large bank of known words that they can automatically recall. Independent spellers continue to use personal constructions when spelling unfamiliar words in draft writing. Independent spellers realise the importance of proof reading.
4 Transitional Spelling	In this phase writers are moving away from heavy reliance on the phonetic strategy towards the use of visual and meaning-based strategies. They may still have difficulty recognising if a word 'looks right' but should be able to proof their known bank of words. Writing will show evidence of an increasing bank of learned words. To help writers at this point it is better not to emphasise phonics but to extend their repertoire of alternative strategies. This is a critical phase in the development of spelling. It often takes writers a long time to move through it. It is important that progress is carefully monitored so as much support and explicit teaching can be given as possible. If writers do not receive sufficient support they may not progress beyond this phase.
3 Phonetic Spelling	In this phase writers are able to provide an almost perfect match between letters and sounds. Letters are chosen on the basis of sound, often without regard for conventional letter patterns. Spelling attempts are meaningful and are becoming more like standard spelling. There is often evidence of self-constructed rules that may not conform to adult rules. Writers copy, recall and construct words according to their current understandings. They use rote recall for an increasing number of words.
2 Semi-Phonetic Spelling	In this phase children show developing understanding of sound–symbol relationships. Their spelling attempts show some evidence of sound–symbol correspondence. They may represent a whole word with one, two or three letters. In this, as in all phases of development, children will be copying, recalling and inventing words. Children at this phase are able to copy letter by letter.
1 Preliminary Spelling	In this phase children become aware that print carries a message. They experiment with writing-like symbols as they try to represent written language. Their writing is not readable by others as understandings of sound–symbol relationships have yet to develop. Children are fascinated by print and are constantly trying to explore the relationships between written and spoken words and between letters and sounds through emulating adults in role play of reading and writing.

(Education Department of Western Australia, 1994, p. i)

The ICT Literacy proficiency scale in Figure 3.6 was developed as part of the Australian National Assessment Program. The six proficiency levels were developed from a statistical analysis of Year 6 and Year 10 students' response to sets of ICT Literacy tasks. The descriptions of the levels include skills and knowledge in information search and evaluation, software

application functions and features, and elements of appropriate and ethical use of ICT (Ainley, Fraillon, Gebhardt & Schulz, 2012, p. 27). The boundary between Levels 2 and 3 was set as the 'proficiency standard' for Year 6 students, and the boundary between Levels 3 and 4, the 'proficiency standard' for Year 10 students.

Figure 3.6: ICT Literacy Proficiency Levels (National Assessment Program)

Level	Proficiency level description
6	Students create information products that show evidence of technical proficiency, and careful planning and review. They use software features to organise information and to synthesise and represent data as integrated complete information products. They design information products consistent with the conventions of specific communication modes and audiences and use available software features to enhance the communicative effect of their work.
5	Students evaluate the credibility of information from electronic sources and select the most relevant information to use for a specific communicative purpose. They create information products that show evidence of planning and technical competence. They use software features to reshape and present information graphically consistent with presentation conventions. They design information products that combine different elements and accurately represent their source data. They use available software features to enhance the appearance of their information products.
4	Students generate well targeted searches for electronic information sources and select relevant information from within sources to meet a specific purpose. They create information products with simple linear structures and use software commands to edit and reformat information products in ways that demonstrate some consideration of audience and communicative purpose. They recognise situations in which ICT misuse may occur and explain how specific protocols can prevent this.
3	Students generate simple general search questions and select the best information source to meet a specific purpose. They retrieve information from given electronic sources to answer specific, concrete questions. They assemble information in a provided simple linear order to create information products. They use conventionally recognised software commands to edit and reformat information products. They recognise common examples in which ICT misuse may occur and suggest ways of avoiding them.
2	Students locate simple, explicit information from within a given electronic source. They add content to and make simple changes to existing information products when instructed. They edit information products to create products that show limited consistency of design and information management. They recognise and identify basic ICT electronic security and health and safety usage issues and practices.
1	Students perform basic tasks using computers and software. They implement the most commonly used file management and software commands when instructed. They recognise the most commonly used ICT terminology and functions.

(Ainley et al., pp. xviii–xix)

Considered together, the examples in Figures 2.8, 2.9, 3.5 and 3.6 illustrate that a learning progression can be based on domains that are defined very narrowly (for example conceptual understanding of energy) or very broadly (for example ICT literacy). Whether defined narrowly or broadly, maps of learning progress have the following characteristics.

- *They describe long-term progress in an aspect of learning.*
  Each empirically based map describes the direction and nature of progress within an area of learning. In general, a learning progression describes increasing degrees of knowledge, understanding and/or skill, but may also describe increasingly positive attitudes, values or dispositions. A learning progression is generally not specific to an age or year level, but describes progress across multiple years of learning.
- *They represent continua of learning.*
  Most learning occurs incrementally over time and can be conceptualised as continuous in nature. Even if a learning progression is divided, somewhat arbitrarily, into a number of 'levels', 'phases' or 'bands', these are simply convenient subdivisions of an underlying learning continuum.

- *They are empirically based.*
  The mapping of learning progress is based on evidence from research and professional experience about how learning typically progresses in that learning area. A learning progression is not based simply on curricular intentions, but on a detailed analysis and study of how learning occurs in practice.
- *They provide frames of reference for assessing and monitoring learning.*
  In describing progress in a domain of learning, an empirically based map provides a frame of reference for assessing where learners are in their learning at the time of assessment and for monitoring progress over time.

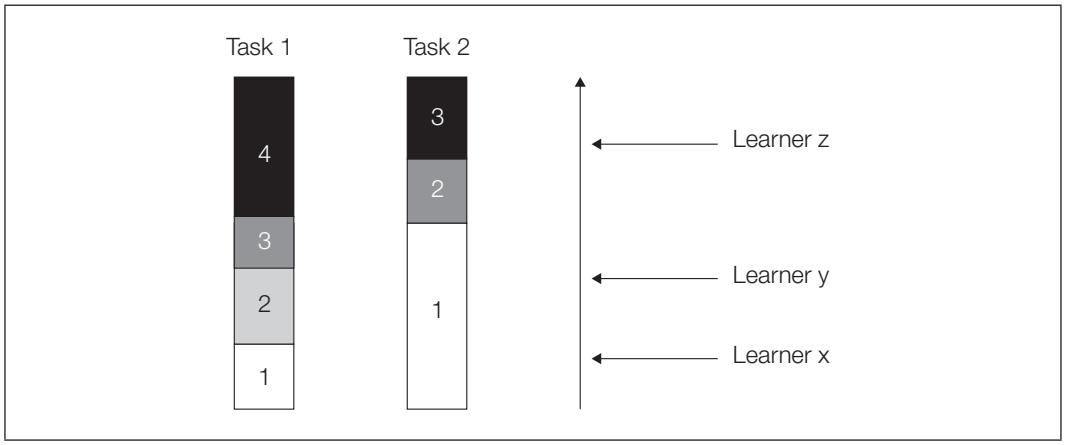
In a Learning Assessment System, conclusions about where learners are in their learning are based on records of their responses to relevant assessment tasks. In this process, task rubrics provide the crucial link between observed responses/performances and the described learning domain. The process is in essence a process of *inferring* where students are in their learning within a domain, and these inferences are always on-balance decisions, based on the available evidence.

There are two general approaches to reaching on-balance conclusions about where learners are in their learning, based on their responses to, or performances on, a set of assessment tasks. These are measurement-based approaches and judgement-based approaches. Both approaches can be employed in the implementation of a Learning Assessment System of the kind described in this research review.

Measurement-based conclusions

The first approach to drawing conclusions about where learners are in their learning is based on the construction of a 'proficiency scale' that describes progress (increasing proficiency) within the domain. Figures 2.9 and 3.6 are examples of proficiency scales constructed in this way. This measurement-based approach uses an individual's responses to a set of assessment tasks to draw an on-balance statistical inference about where they are on the proficiency scale constructed for the domain. The process can be illustrated using the hypothetical example in Figure 3.7.

Figure 3.7: Assessment task rubrics calibrated on a proficicency scale



(Masters, data file)

The vertical arrow in Figure 3.7 represents a continuum of increasing proficiency within a particular learning domain. Along this continuum, the estimated locations of three individuals (x, y and z) are shown. These estimated locations are statistical estimates based on each individual's responses to a set of domain-appropriate assessment tasks. They are on-balance estimates that take into account all the available evidence about each learner from his or her performances on the set of tasks. Learner z is estimated to be at a higher level of proficiency within this domain than learner y, who in turn, is estimated to be at a higher level of proficiency than learner x.

The measurement process simultaneously positions ('calibrates') each of the tasks in this set of tasks on the same proficiency continuum. On the left of the figure, two tasks from the full

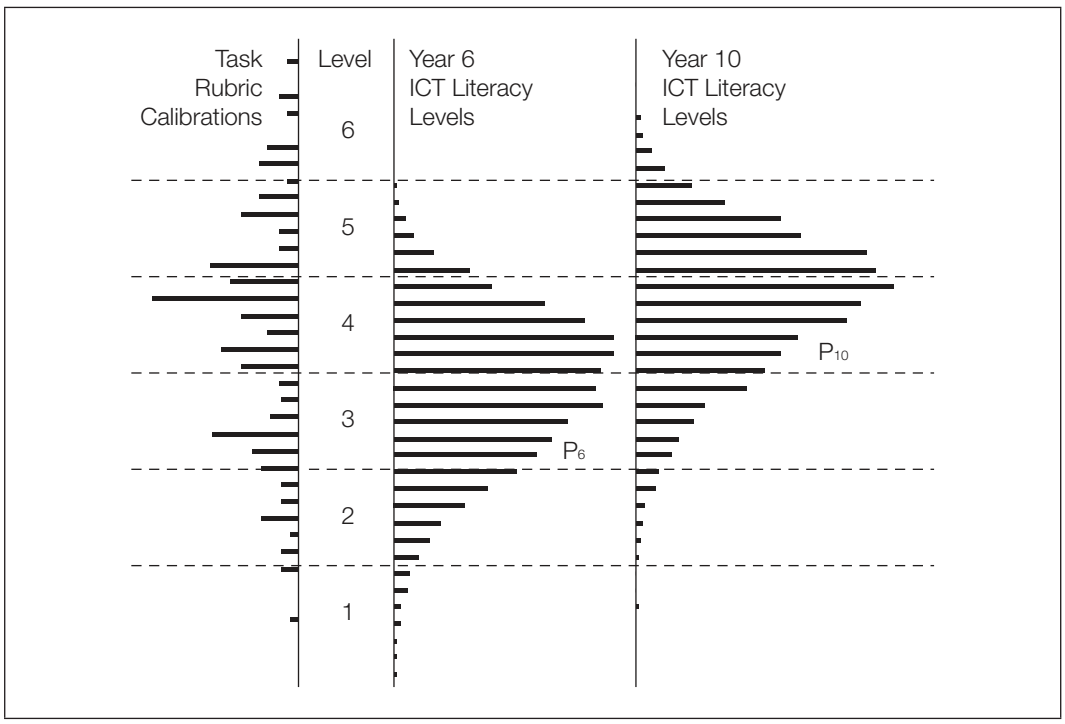
set of tasks positioned on this proficiency continuum are shown. The shaded regions correspond to the levels of the rubric for each task. The rubric for Task 1 had four levels; the rubric for Task 2 had three. It can be seen that level 2 on Task 2 represents a higher level of proficiency than level 2 on Task 1.

As discussed earlier in Section 3, the levels of these two task rubrics ideally would have been developed originally from an understanding of the domain itself. For each task, the hierarchy of levels describes – and is a task-specific illustration of – increasing proficiency within the domain. Once statistically positioned on a proficiency scale, the levels of the task rubrics assist in giving substantive meaning to locations along the scale. For example, the region of the proficiency scale in which learner y is located would be described in part by reference to the descriptions of level 2 of Task 1 and level 1 of Task 2. The described proficiency scales in Figures 2.9 and 3.6 were developed from exactly this kind of analysis.

This process of calibrating task rubrics on a proficiency scale and developing on-balance estimates of learners' locations on this same scale is the purpose of modern educational measurement. Psychometrics is sometimes considered to relate only to tests, but this is a misconception. The main focus of modern measurement theory is on drawing valid inferences from assembled assessment evidence, and psychometric methods can be applied to any set of assessment tasks for which rubrics have been developed. Nevertheless, it is also true that the major international assessment programs (PISA, TIMSS, PIRLS), the Australian National Assessment Program (NAPLAN, Science, Civics and Citizenship, ICT Literacy) and many commercial testing programs (for example, PAT Progressive Achievement Tests, TORCH Tests of Reading Comprehension) use modern measurement theory to develop and report against described proficiency scales.

An example of the use of measurement theory to establish where learners are in their progress within a learning domain is shown in Figure 3.8. This example shows the details used in the construction of the described proficiency scale in ICT Literacy in Figure 3.6.

Figure 3.8: Construction of an ICT literacy proficiency scale



(Masters' data file, based on Ainley et al., 2012, p. 26)

On the left of Figure 3.8, ICT Literacy tasks have been calibrated along a continuum. The easiest tasks are calibrated at the bottom of the scale; the most difficult tasks are calibrated at the top. The scale has been divided into six proficiency levels and the tasks located in each of these levels have been used to develop the level descriptions in Figure 3.6.

On the right of Figure 3.8, the distributions of Australian Year 6 and Year 10 students' estimated locations on this ICT Literacy scale are shown. The Year 6 and Year 10 proficiency standards (P_6 and P_{10}) also are marked. As might be expected, Year 10 students are generally at more advanced levels of ICT Literacy proficiency than Year 6 students. However, as observed earlier for other learning domains, there is significant overlap in these distributions.

Because this first approach to inferring where learners are in their learning underpins many widely used classroom assessment resources and is the basis of assessments made using national and international assessments such as NAPLAN, PISA, TIMSS and PIRLS, it is important that teachers and school leaders have at least a working knowledge of this approach and its implementation. The advantage of the measurement-based approach, when applied to professionally constructed assessment tasks, is that it is capable of providing teachers, schools and education systems with valid, reliable and objective assessments of student progress in aspects of their learning.

However, it is generally not practicable for classroom teachers to apply this measurement-based approach to their day-to-day classroom assessments. Outside the context of professionally developed assessment resources, teachers must usually use their professional judgements to decide where individuals are in their learning.

Judgement-based conclusions

An alternative to measurement-based conclusions is professional judgement. In this process, assessors (usually teachers) compare the observations they have made with the descriptions of the levels on a proficiency scale. The process is one of finding the best match by weighing the available evidence mentally and making a global, on-balance judgement.

> *In holistic or global grading, the teacher responds to a student's work as a whole, then directly maps its quality to a notional point on the grade scale. Although the teacher may note specific features that stand out while appraising, arriving directly at a global judgment is foremost … Determining the quality of divergent types of works requires skilled, qualitative judgments using multiple criteria. A qualitative judgment is one made directly by a person, the person's brain being both the source and the instrument for appraisal.*

(Sadler, 2009, p. 1)

For judgements of this kind to be meaningful, the learning progression needs to be made explicit and, ideally, be illustrated with examples of the kinds of knowledge, skills and understandings typically demonstrated by students at varying points along the progression. An example of judgement-based conclusions of this kind is teachers' use of the *First Steps* spelling developmental continuum (Figure 3.5). In addition to general descriptions of the five phases of this continuum, a number of accompanying indicators and examples of spelling proficiency at each phase are provided. Students are said to be 'working in' a particular *First Steps* phase when the global statement and key indicators represent the best fit with the student's learning behaviours. 'Children need not display all key indicators to be placed in a phase; the placement rests upon the teacher's professional judgement' (Education Department of Western Australia, 1997).

Sadler (2009) describes several advantages of holistic judgements. These advantages include the fact that holistic judgements depend on assessors developing an understanding of the big picture of learning progress within a domain. For example, a teacher using the *First Steps* spelling continuum can make an on-balance judgement of an individual's level of spelling development only by engaging with the full set of phases. This general understanding of the nature of progress can be useful to learners themselves, particularly older learners:

> *If students are to achieve consistently high levels of performance, they need to develop a conceptualisation of what constitutes 'quality' as a generalised attribute.*

(Sadler, 2009, p. 2)

Sadler also argues that holistic, on-balance judgements are capable of providing more valid estimates of student achievement than analytical approaches because some important aspects of student achievement may not be identified in a pre-specified set of criteria and because overall judgements may be more than the sum of their parts.

In summary

The purpose of a Learning Assessment System is to draw a conclusion about where learners are in an area or domain of learning; this domain can be broadly defined or very specific. Assessment conclusions are based on responses to relevant assessment tasks. The process requires a deep understanding of the learning domain itself and is essentially a process of drawing an inference from observed responses/performances. Qualitatively described task rubrics provide the crucial link between task performances and the learning domain. Conclusions are on-balance decisions which can be made either statistically with the assistance of a measurement model or impressionistically as global, holistic judgements.

Feedback and monitoring

The outcome of the learning assessment process is a conclusion about where learners are in an aspect of their learning at the time of assessment. This conclusion may be for an individual learner (for example, a conclusion about a student's current level of spelling proficiency) or for a group of learners (for example, a conclusion about the average scientific literacy level of Australian 15-year-olds). Assessment conclusions inform starting points for action by teachers, students, parents, school leaders, system leaders and governments and also provide a basis for monitoring learning progress and evaluating the effectiveness of past actions.

When used in these ways, assessment becomes an integral part of an ongoing educational decision-making process. It is a way of continually checking on current levels of learning and achievement and of monitoring progress over time. At the classroom level, rather than being a process that is separated from and comes after teaching and learning, assessment becomes an essential component of effective pedagogy and a vital part of successful learning.

A learning culture

Assessment is an important element in establishing and maintaining a learning culture – within classrooms, schools and across entire education systems. A learning culture is one in which there is a deep belief in the possibility of continual improvement; a shared commitment to understanding the status quo and to monitoring progress; a strong desire for information about effective ways of improving on current levels of achievement; an openness to self-reflection and the critical evaluation of current practice; and a willingness to make changes required for improvement.

As noted in Section 2 of this review paper, the conditions for successful learning, including successful teacher and leader learning, include positive emotional engagement, intrinsic motivation, a sense of safety and support to take risks, strong self-efficacy as a learner, confidence that effort will lead to success, and metacognitive (self-monitoring) skills. Negative emotions such as stress and fear of failure inhibit successful learning.

Assessment and reporting processes can both promote and undermine cultures of learning. When assessments are focused only on judging student success, they encourage *performance cultures* rather than *learning cultures* in classrooms. One common way of judging student success is to compare student performances with course or year level expectations and to assign grades (usually A to E) to indicate the extent to which those expectations have been met. However, as was established in Section 2, students in the same year of school can be at very different stages in their learning (see Figures 2.2 to 2.4), making it inevitable that some students will not meet year level expectations. These students may be judged to be 'failing', or at least underperforming, and awarded low grades, possibly year after year. On the other hand, more advanced students

may have little difficulty in meeting year level expectations, and so achieve A grades year after year, without being appropriately challenged and extended.

Another common way of judging student success is in terms of the performances of other students. In some educational contexts, students compete for high rankings or high grades. For example, the practice of 'grading on the curve' assigns grades to fixed percentages of students, guaranteeing that some students will receive low grades regardless of their absolute levels of achievement. When learning is driven by extrinsic factors such as competition for high grades or fear of failure, the quality of learning itself often suffers.

Assessment and reporting processes are more likely to promote classroom learning cultures when they are based on an understanding by all participants that learners can be at very different points in their learning and a belief that, despite this, all learners are capable of learning progress if motivated and if provided with appropriate learning opportunities. From this perspective, the role of assessment is to establish where individuals are in their learning so that teaching can be differentiated and further learning progress can be monitored over time.

Traditional forms of assessment in which every student undertakes the same set of tasks may not be the most effective way to establish where individuals are in their learning, unless those tasks are sufficiently open-ended that they allow students at different levels of achievement to engage successfully and experience a degree of success. Alternative forms of assessment such as multi-level assessments and computer adaptive tests in which students undertake tasks tailored to their current levels of achievement are more appropriate when learners are at very different points in their learning.

Supporting self-monitoring

Assessment processes also can promote classroom learning cultures by encouraging and supporting student involvement in the monitoring of their own learning. When students are encouraged to reflect on their learning they are better able to set goals for future learning, to take a degree of responsibility for their own learning, and to monitor the learning progress they are making.

In traditional approaches to teaching, learning and assessment, teachers deliver a curriculum that students are expected to learn. Assessments are undertaken at various points during teaching and learning to establish how well students have learnt what has been taught, and the outcome of the assessment process is a judgement of learning success. Self-assessments and self-monitoring can occur within this traditional paradigm, with students asking questions about and judging their own success. How much of the taught content have they learnt? What are the gaps in their learning? What grade would they assign to their performance? Peer assessments also can take the form of students judging each other.

In a Learning Assessment System of the kind advocated in this review, self-monitoring is undertaken in relation to an explicit description of long-term learning progress within a learning domain. Learners are supported to view learning as progress towards higher levels of knowledge, skill and understanding, to reflect on where they are in their learning, and to set goals for further learning progress. Central to self-monitoring is a scientifically based understanding of the learning domain and, in particular, an understanding of what it means to make progress (for example, the characteristics of high-quality work and performances).

Feedback from assessments is essential to self-monitoring. Feedback is most effective when it assists learners to self-monitor and provides clear guidance on specific actions that can be taken to make further learning progress. Feedback needs to be timely, to assist students to see where they are currently in the learning, to be in a form that clarifies the progress that students make over time, and to be provided in ways that promote a belief that further learning progress is possible (Hattie, 2003, 2009).

In practice, feedback from assessments is often not provided in forms that enable this level of self-monitoring. If the only feedback students receive from an assessment process is a grade, percentage or uninterpreted score, then learners are unable to see where they are in their long-term learning and virtually no guidance is provided on the steps students can take next in their learning.

When learners are assisted to see where they are in their long-term progress within a learning domain, they are better able to set specific goals for their future learning. Rather than teachers specifying a common set of learning intentions for all students, learning goals and plans can be jointly personalised to individuals' current levels of achievement and learning needs. In supporting students to set learning goals and to monitor their own progress, teachers encourage the development of metacognitive skills and assist learners to take a level of responsibility for, and control over, their own learning.

Self-monitoring of this kind also has the ability to build learners' self-confidence in their own ability to learn successfully. Indeed, there may be no better way to build a learner's belief in their ability to learn than to enable them to see the learning progress they make over time. As noted earlier, common reporting methods, such as grades, do not usually communicate learning progress and can undermine individuals' beliefs in their capacity for successful learning.

In contrast, when learning is monitored against an explicit description of long-term progress within a domain, feedback to students and parents can be provided in forms that make clear the progress that is being made. As well as describing this progress, it can be illustrated with samples of individuals' work and performances. For example, samples of a student's writing, work in Mathematics, or video-recordings of their oral reading made over months or years might be used to demonstrate clearly to learners and to their parents/carers the progress that is being made. Again, the purpose is not so much to judge learners as to understand and communicate the learning that is occurring and that it can continue to occur.

Finally, the monitoring of learning progress provides an alternative way of thinking about 'successful' learning. Under traditional approaches, successful learning is defined only as the degree to which the taught curriculum has been learnt and the course or year level expectations have been met. This review is proposing that successful learning be defined as *excellent progress*. Under this definition, a student would be considered to have learnt successfully if they made excellent progress, regardless of where their absolute level of achievement was in relation to year level expectations. There may still be a place for assessing students against expectations for their age or year level, but rather than setting learning expectations *only* in terms of year level standards, expectations can also be set for the progress that all students should make in their learning (for example, over the course of a school year). Such a definition of success would enable less advanced students to monitor and celebrate successful learning progress.

In summary

Step 4 of the Learning Assessment System described in this review provides feedback that assists learners and other interested parties to see where learners are in their long-term progress within a learning domain and to monitor progress over time. In this way, assessment becomes an integral part of professional decision-making and is an essential feature of learning itself. Assessments are based on the belief that, while learners are at different points in their learning and may be progressing at different rates, all learners are capable of successful learning if motivated and provided with appropriate learning opportunities. As a result, assessments become part of a learning culture, with learners being assisted to set goals and plans for their learning and to monitor the progress they make over time.

Concluding comments

This section has described a set of principles and steps for an assessment system that is based less on judging how well students have learnt what they have been taught than on establishing and understanding where students are in their long-term learning progress within a learning domain. This distinction between assessment as judging and assessment as understanding is at the heart of a Learning Assessment System and is summarised in Figure 3.9. The figure is organised into the five key steps in a Learning Assessment System, as described in this section.

Figure 3.9: Distinctions between assessment as judging and assessment as understanding

The learning domain	
Assessment as judging • a specification of what teachers are to teach and students are to learn, usually in the form of a curriculum or course syllabus	***Assessment as understanding*** • a description of the nature of learning within a defined area or domain of learning
• typically developed by a curriculum agency, informed to varying degrees by research	• strongly grounded in learning research (i.e., scientifically based)
• the focus is on content to be taught and outcomes to be achieved – largely from the perspective of teachers	• the focus is on understanding and describing the nature of learning – largely from the perspective of learners
• includes the specification of knowledge, skills and understandings, as well as higher-order skills and attributes that make up the domain	• includes the identification of typical learning sequences/progressions; prerequisites for successful learning; learning difficulties and obstacles; common student errors and misunderstandings
• the emphasis tends to be on specifying the horizontal (content) structure of the domain	• there is a strong emphasis on describing the nature of long-term learning progress within the domain (i.e., vertical as well as horizontal structure)
• the focus on content coverage often results in overcrowding of the curriculum and an overemphasis on relatively superficial learning (e.g., factual and procedural knowledge resulting in 'mile-wide, inch-deep' curricula)	• a consequence of focusing on the vertical structure of the domain is a greater emphasis on attributes that develop only over extended time periods (e.g., deep understandings; higher-order thinking; life skills and attributes)
• the domain is commonly specific to a particular course, unit or module to be delivered over a specified time period (e.g., a particular year of school)	• the domain is a description of long-term learning progress (i.e., not time-limited and usually not grade-specific)
• takes the form of a catalogue of content to be taught and outcomes to be achieved	• takes the form of a complex, empirically based 'map' of how learning occurs within a domain
Assessment methods	
Assessment as judging • it is generally assumed that there are multiple purposes for assessment and that those multiple purposes require quite different methods of assessment • despite this, most assessments are designed for the general purpose of establishing whether or not students have learnt what they have been taught (and to report how much of what they have been taught students have successfully learnt)	***Assessment as understanding*** • assessments are undertaken for a single general purpose: to establish where learners are in their learning within a domain at the time of assessment • assessments for this purpose can be undertaken at any time (during a course, at the end of a course, or without reference to a course at all) and the results of the assessment process can be used in a variety of different ways • the timing and proposed uses of assessment results are generally irrelevant in choosing an appropriate assessment method
• it is common to distinguish different forms of assessment, for example, – formative assessments are undertaken during a teaching course to track how well students are learning what is being taught (to inform ongoing teaching) – summative assessments are undertaken at the end of a course to establish how well students learnt what was taught in the course – diagnostic assessments commonly are undertaken to identify gaps in student learning (i.e., taught content that has not yet been learnt)	• many traditional assessment distinctions are less relevant. For example, the same information about where students are in their learning can be used both prospectively to inform future teaching (assessment for learning) and retrospectively to evaluate learning progress (assessment of learning). Assessments of where students are in their learning are not fundamentally different whether undertaken during or at the end of a course. Diagnostic assessments are designed to understand in greater detail where students are in their learning, particularly by identifying specific difficulties, faulty strategies, misconceptions and ways of thinking

• choices of assessment methods often are based on personal preferences and beliefs about what will be valid and fair in judging student learning success. For example, there may be generalised preferences for school-based assessments over externally developed assessment tasks; 'authentic' assessment tasks over devised assessment tasks; or 'performance' assessments over standardised multiple-choice tests	• methods of assessment are chosen primarily for their capacity to provide (construct) valid information about where students are in their learning within a domain. • no assessment method is considered inherently superior to any other; methods valid for some aspects of learning will not be valid for others. Construct validity is an empirical matter, not a matter of personal preference
Recording responses to tasks	
Assessment as judging	***Assessment as understanding***
• responses to assessment tasks provide the evidence for judging how well students have learnt what they have been taught	• responses to assessment tasks provide the evidence for establishing where learners are in their long-term learning progress within a domain
• various formats are used for recording students' responses to (or performances on) assessment tasks	• responses to each task are recorded in terms of a 'rubric' consisting of two or more qualitatively described, ordered levels of response to (or performance on) that task
• formats for recording responses to assessment tasks are not explicitly aligned with the learning domain. For example, individual pieces of student work, such as a project, essay, assignment or test, may be graded (e.g., A to E) with no accompanying grade interpretations; or task rubrics may consist of little more than a general description qualified at each level by graded adjectives such as 'limited', 'sound', and 'excellent'	• the levels of a rubric describe observable responses or performances (that is, task-specific examples of the learning domain). These typically are developed from, can be mapped on to, and also may help to inform the structure of, the learning domain. For example, the rubric for a test item can contribute to the construction of a proficiency scale and so assist in describing and elucidating the nature of learning progress within that domain
Drawing assessment conclusions	
Assessment as judging	***Assessment as understanding***
• conclusions are judgements about how well students have learnt what they have been taught	• conclusions are drawn about where learners are in their learning progress in a domain at the time of assessment. Conclusions may relate to progress in specific sub-areas of learning, prerequisite skills and understandings, or learners' difficulties, errors and misconceptions
• conclusions usually are based on simple aggregations of students' performances on a common set of assessment tasks	• conclusions are based on recorded responses to or performances on assessments tasks. Because assessment task rubrics are mapped to the learning domain, inferences about where learners are in their learning can be based on different tasks
• assessments made at various times during a course (i.e., 'continuous' assessments) may be brought together and aggregated for this purpose	• assessments can indicate where learners are in a particular aspect of their learning at different times, however the aggregation of assessments made at different times is considered largely meaningless
• conclusions take the form of course grades (e.g., A to E) or percentages – both designed to convey how much of the course content students have successfully learnt; course failure is a possibility	• conclusions take the form of inferences about where learners are in their progress through a learning domain, conveyed descriptively, pictorially and/or numerically (e.g., location on a proficiency scale)
Feedback and monitoring	
Assessment as judging	***Assessment as understanding***
• feedback is focused on how well students have learnt what they have been taught; this may include the identification of gaps and areas requiring further attention	• feedback is focused on where learners are in their long-term learning, what progress they are making, and what specific actions they can take next in their learning
• feedback usually includes a report on learning success, typically in the form of a course grade or score; percentiles may be used to compare students' performances with the performances of other students	• learning success is defined and reported as progress made; the focus is on individual learning journeys, although comparisons with age/grade norms and/or expectations may be provided
• feedback sometimes is provided in forms that fail to reveal learning progress and reinforce low self-perceptions as learners.	• feedback is designed to show long-term learning progress and to encourage a belief that successful learning is possible.

(Masters, data file)

The two columns of this figure are perhaps best thought of as points towards opposite ends of a continuum. Although assessment in education continues to be dominated by a focus on judging student success, much current practice combines elements of these two approaches. As argued throughout this review paper, the goal of assessment reform should be to achieve a greater focus on the approaches and processes on the right of this figure; to focus less on assessment as judging, and more on assessment as understanding.

Section 4 of this review builds on the previous sections by identifying four general challenges to achieving assessment reform, and considers some implications for assessment policy development.

section

4

Achieving assessment reform

This review has considered the case for reforming the field of educational assessment and has argued that significant reform is needed. Section 2 reviewed a number of current pressures for reform and sketched some of the ways in which traditional approaches to assessment and reporting are out of step with the present information needs of learners and those who work to support them. To be more useful at all levels of decision-making, educational assessments must be more strongly grounded in scientifically based understandings of learning, particularly research into how learning occurs within specific domains. It was noted that traditional approaches to assessment based on judging how well students have learnt what they have been taught are largely inappropriate for assessing and monitoring the development of deep understandings of subject matter and life skills and attributes that develop over extended periods of time. And it was observed that assessments designed to support traditional modes of educational delivery are increasingly inappropriate, due to advances in technology transforming where and when learning takes place, enabling more personalised forms of learning and assessment.

Section 3 outlined a set of design principles for a Learning Assessment System in which the purpose of assessment is not so much to judge student success as to understand where learners are in their long-term learning at the time of assessment and to monitor learning progress over time. This final section outlines some practical challenges and some ways forward in reforming educational assessment along the lines described in this review.

Repurposing assessment

A first general challenge arises from the widely held perception that the fundamental purpose of assessment in education is to judge how well students have learnt what they have been taught. Much of the field of assessment, including many assessment concepts and much of the language of assessment, was developed from this perspective.

Over time, attempts have been made to move beyond this traditional conception of assessment. For example, following the introduction of the concepts of formative and summative program evaluation (that is, the evaluation of educational programs in the course of their delivery as well as upon completion), the concepts of formative and summative student assessment were introduced to encourage assessments not only upon completion of a course of instruction, but also during course delivery to inform teaching and learning (Bloom, 1968). In classroom practice, however, formative and summative assessments often differ only in their timing and are undertaken within the same general paradigm of judging how well students have learnt what they have been taught.

Other attempts at assessment reform tend to have divided the field into multiple purposes, philosophies and methods, each with its own protagonists. Rather than reforming the field, these efforts often have produced sub-fields, which usually correspond to the specific methods and approaches being promoted by their proponents as inherently more desirable than others.

This review paper has argued for reconceptualising the essential purpose of assessment. In particular, it has argued that assessments should be seen as having a single general purpose: to establish where learners are in their long-term progress within a domain of learning at the time of assessment. The purpose is not so much to judge as to understand. This unifying principle, which has potential benefits for learners, teachers and other educational decision-makers, can be applied to assessments at all levels of decision-making, from classrooms to cabinet rooms. Generally, the same assessments will be useful both for monitoring the progress that individuals or groups are making over time (that is, assessments of learning) and for identifying starting points for future action (that is, assessments for learning). More detailed classroom diagnostic assessments will sometimes be required to achieve a more complete understanding of where learners are in their learning, for example, by exploring students' specific difficulties and misunderstandings.

Although seemingly simple, this reconceptualisation represents a paradigm shift; one which, it has been argued through this review paper, has significant implications for practice. Most assessment practice is intimately linked to the current model of educational delivery, and any change in assessment practice is likely to impact on educational delivery and vice versa.

Assessment is an integral part of the prevailing model. Its role is to establish how well students have learnt (or in the case of formative assessment, how well they are learning) what teachers have taught. The prevailing view of assessment as judging student success is deeply embedded not only in educational practice, but in society more generally. The grading of student success is ubiquitous – so much so that attempts to reform educational assessment have often accepted summative grading as a given and either assigned it a legitimate place alongside other more 'desirable' and teaching-oriented forms of assessment or argued for particular forms of assessment (for example, holistic teacher judgements) as the basis for student grades. The simplicity of A to E grades and the (generally erroneous) belief that they convey meaningful information about learning progress have contributed to this situation.

In the 21st century, assessments designed only to judge student success against the performances of other students or against age/grade expectations are no longer adequate. The new purpose of assessment requires practitioners and learners to begin with the belief that every learner is capable of excellent learning progress, whatever their current starting point. If learning is to be judged, then it is more appropriate that judgements are based on the progress that individuals make in their learning than on their ability to demonstrate year level expectations (although there will continue to be some value in monitoring performances against such expectations).

However, assessment reform is likely to be difficult in the absence of broader educational reforms. For example, assessment to establish where students are in their learning is largely pointless if teachers intend to deliver exactly the same content to all students in a class regardless of their current levels of achievement, if the overriding assessment concern is the generation of grades which convey how well students have mastered the curriculum for their year level, or if governments demand graded judgements of student learning. On the other hand, the reform of assessment thinking and practice has the potential to lead and drive improvements in teaching and learning.

Perhaps the most significant challenge in reforming assessment along the lines described in this review paper is that it requires a change in mindset, particularly on the part of stakeholders who consider it 'equitable' to hold all students of the same age to the same absolute achievement standards. Accepting the reality that students in the same year of school are at very different points in their learning and so are likely to benefit from differentiated teaching, different learning targets and different measures of learning success is not a matter of accepting lower expectations for some students' learning. If expectations are couched in terms of student *progress*, then there is no argument or reason why the same high expectations of progress should not be set for all

students, regardless of their absolute levels of achievement. Failure to identify an individual's learning needs and attempting instead to infer those needs from group membership (for example, age or year level) is generally likely to be more 'inequitable' and more detrimental to learning.

From the perspective of teachers, the challenge will hinge on embracing the implications of personalised learning for assessment. Personalisation of learning implies using assessment to establish where individuals are in their learning, setting personal learning goals, providing differentiated learning opportunities, monitoring individual learning progress, and encouraging self-monitoring. As previously noted, personalisation and differentiation are undermined by assessment and reporting practices that fail to recognise individual progress and that judge success or failure only in terms of year level standards/expectations.

Arguments for traditional forms of reporting are sometimes made on the grounds that parents understand grades and percentages. However, in the main, parents do not understand traditional forms of reporting because most grades and percentages lack consistent meanings across teachers, subjects and schools. Regression to these traditional report formats sometimes occurs in response to well-intentioned, but unsuccessful and overly complex, attempts to provide parents with better descriptive information. A long-term educational challenge is to develop alternative reporting formats that provide parents and carers with more usable information about where individual learners are in their ongoing learning, what progress they have made, and what might be done to support further learning.

Education systems and governments are in strong positions to influence perceptions of assessment. On one hand, they can promote a traditional view of teaching as the delivery of a common curriculum to all students in the same year level, assessment as the process of establishing how much of this common curriculum each student has learnt, and reporting as the grading and communication of student success. Or they can promote a view of teaching as the process of identifying and addressing the learning needs of individual learners, assessment as establishing where individuals are in their learning, and reporting as the communication of information about learning progress. Although there is a place for explicit year level expectations, it is incumbent on education systems and governments to promote practices that do not define success *only* as the achievement of year level expectations. Such traditional practices run the risk that less advanced learners will be viewed (and will view themselves) as 'poor' learners and that the learning needs of these students will not be identified and addressed. There is a parallel risk that, because year-level expectations are relatively easily met by more advanced learners, the learning needs of these students also will be inadequately identified and they will remain unchallenged. The starting point for assessment and reporting policy must be a belief that all students can and should make excellent learning progress and that the key purpose of assessment is to establish where learners are in their learning in order to promote further learning progress.

Mapping learning domains

A second general challenge is the development of carefully constructed, research-based understandings of how learning proceeds within particular domains of learning. The concept of an empirically based learning domain is at the heart of the Learning Assessment System described and advocated in this review paper. Indeed, the essential purpose of assessment in such a system is to establish where students are in their learning progress within well-understood and described learning domains.

An empirically based learning domain is related to, but also differs from, most syllabus or curriculum specifications in two important respects. First, most school curricula are specifications of year-level content 'standards' or expectations. They specify, often in broad terms, the standards expected of students by the completion of each year of school.

Figure 4.1, for example, shows expectations specified in the Australian Year 3 English Curriculum. These expectations relate specifically to the interpretation, analysis and evaluation of texts. The changed expectations from the corresponding portion of the Year 2 English curriculum

are shown. In particular, by Year 3, students are expected to identify not only the audience but also the purpose for imaginative, informative and persuasive texts; they are expected to read a broader range of text types; to use the additional reading strategies of confirming and reading on; and they are expected to move beyond the analysis of texts to their evaluation.

Figure 4.1: Example of year level curriculum expectations

*Year 3 English (interpreting, analysing, evaluating)**

1 Identify the audience and purpose of imaginative, informative and persuasive texts
2 Read ~~*less predictable*~~ *an increasing range of different types of texts* by combining contextual, semantic, grammatical and phonic knowledge, using text processing strategies, for example monitoring, predicting, *confirming*, rereading, *reading on* and self-correcting
3 Use comprehensive strategies to build literal and inferred meaning and begin to ~~analyse~~ *evaluate* texts by drawing on growing knowledge of context, language *features and text structures* ~~and visual features and print and multimodal text structures~~

*changes from Year 2 curriculum expectations shown by strikethroughs and italics

(Australian Curriculum, Assessment and Reporting Authority, 2012, p. 24)

These curriculum expectations provide a useful guide to some general skills that students should demonstrate by the end of Year 3. But they provide a limited guide to *teaching* Reading to Year 3 students. As noted in Section 2, students in Year 3 differ in their reading levels by as much as five or six years of school. If Year 3 teachers were to interpret their role simply as 'delivering' these elements of the Year 3 English curriculum, then they might be tempted to focus their teaching of Reading on ensuring that all students were able to identify the purposes of texts, use the reading strategies of 'confirming' and 'reading on', and so on. To adopt such an approach to teaching would fail to recognise the existing variability in students' reading levels and would demonstrate a misinterpretation of the purpose of general curriculum standards.

Broad curriculum expectations of the kind shown in Figure 4.1 also provide a very limited guide to assessing where individual learners are in their reading progress or what difficulties they might be experiencing. Some Years 3 students will still be at very early stages of reading development and have limited foundational, pre-reading skills. To establish where individuals are in their reading progress, teachers require a detailed, research-based understanding of how reading proficiency develops over extended periods of time, including an understanding of common reading skill deficits and reading difficulties. In other words, teachers need access to a research-based map of the Reading domain.

A key characteristic of an empirically based learning domain of the kind described in this review is that it is not linked to specific ages or year levels, but provides a description of long-term learning progress within a domain. It recognises that learners of the same age and year level can be at very different points in their learning.

A second key characteristic on an empirically based learning domain is that it goes beyond a specification of general curriculum expectations or content standards, in that it attempts to describe *how* learning occurs within the domain. There is a strong focus on describing the nature of long-term learning progress, generally identifying:

- common learning sequences and progressions
- prerequisites (knowledge and skills) for successful learning
- common obstacles to learning, including learning difficulties, misunderstandings and student errors.

In other words, the domain is heavily based in pedagogical content knowledge about the nature of learning within the domain. Assessments are then undertaken to establish where students

are in their learning within this empirically based understanding of learning (rather than being focused only on the achievement of specified year-level expectations).

The descriptions of an empirically based learning domain are developed from detailed investigations of how learning occurs, usually based on decades of research and professional experience. Questions addressed through such domain-based research include the following:

- What is the nature of developing proficiency within the domain?
- What distinguishes experts from novices?
- What knowledge is required for high-level performance?
- What conceptual frameworks do learners establish to organise their knowledge?
- What is the nature of increasingly deep understanding within the domain?
- Are there common sequences in which learning develops?
- Are these 'natural' learning progressions, rather than the result of curriculum conventions?
- Are there prerequisites for learning progress?
- What misconceptions do learners commonly develop?

Research of this kind results in deep understandings of the domain, which are required to establish where individuals are in their learning and to design appropriate learning experiences.

This second general challenge relates first to the construction of research-based learning domains, and then to the promulgation and use of research-based domains to guide the design of assessments. For example, if learning research demonstrates that, at a given stage of learning, a particular skill deficit or a specific misconception is an impediment to further progress, then assessments will need to be designed to expose those skill deficits and misconceptions. Unfortunately, the fact that much educational assessment has not been strongly guided by learning research has meant that assessment results often have had limited pedagogical value.

Implementing a Learning Assessment System

A third general challenge is how to implement a coherent Learning Assessment System in a range of practical educational assessment contexts. Much assessment practice in education is not based on a clearly identified assessment 'system', underpinned by a set of principles. In fact, assessment is often seen as though it is straightforward, a consequential extension of a curriculum. Once curriculum standards/expectations are clear, assessment is often seen simply as the processes of checking, with reporting being the process of communicating to stakeholders, how well those expectations have been met. A belief that educational assessment is unproblematic may, in part, explain why assessment design receives so little consideration in most curriculum development efforts and so little time and attention in most teacher education programs.

A coherent Learning Assessment System is one in which the elements are mutually supporting and interdependent and work together for a common purpose: to establish where learners are in their progress within a well-defined learning domain. A Learning Assessment System is a common template or framework for designing educational assessment processes for a learning domain. And, because it is based on a set of general design principles, a Learning Assessment System is intended to be equally relevant in all educational assessment contexts, including classroom diagnostic assessments, international surveys, senior secondary assessments, national literacy and numeracy assessments, and higher education admissions testing.

As noted previously, a prerequisite for a Learning Assessment System is a deep, research-based understanding of how learning occurs within a domain, and the purpose of assessment is to establish where learners are in relation to this understanding. In contrast, much assessment practice in education does not commence with research-based understandings of learning. Often, assessment commences with curriculum scope and sequence charts (which themselves might be informed to varying degrees by research). In other cases, assessments lack conceptual frameworks entirely and are developed by reaching consensus among users or are simply based on tasks that are assumed to provide useful information about the domain. A consequence can be a heavy emphasis on assessing factual and procedural knowledge at the expense of assessing deeper understandings. Often the outcome is a limited representation of where students actually

are in their learning. (For example, students can sometimes perform very well in applying formulas to solve examination problems, but retain fundamental misconceptions of key principles within a domain.) In a Learning Assessment System, construct validity means more than alignment with a scope and sequence chart; it means alignment with a research-based understanding of the nature of learning within the domain.

In recent decades, efforts have been made to promote the use of a wider range of educational assessment methods (for example, portfolios of classroom work, rich/complex tasks and computer-based simulations). However, although the goal of broadening the range of assessment methods has been a worthy one, these alternative forms of assessment have generally not been promoted and taken up as part of a coherent assessment *system*. Attempts at assessment reform have often been based simply on the promotion of specific methods, such as 'performance' assessments, teacher assessments, 'authentic' assessments, and interactive electronic assessments.

As noted in Section 2, the wider use of electronic learning environments is introducing the possibility of significantly new forms of electronic assessment. However, here again, such evidence needs to be collected as part of a larger assessment system. Much current electronic assessment is little more than traditional assessment delivered on a screen (for example, banks of tasks to assess whether students can demonstrate what they have been taught), and the greatly enhanced ability to capture student response data is not always being used to draw systematic inferences about student learning. Mislevy, Behrens, Dicerbo and Levy (2012) describe the possibility of using 'educational data mining' to extract assessment information from interactive learning environments. They note that information collected in this way could lead to better understandings of learning domains, but they also note that, for the purposes of assessing where students are in their learning, student responses need to be interpretable in terms of the domain of interest:

> *It is easy to amass rich and voluminous bodies of low-level data, mouse clicks, cursor moves, sense-pad movements, and so on, and choices and actions in simulated environments. Each of these bits of data, however, is bound to the conditions under which it was produced, and does not by itself convey meaning in any larger sense … Just having gigabytes of keystrokes and mouse clicks is not sufficient for claiming one has good evidence for a particular purpose… It is necessary to recognise recurring and substantively salient features of situations, so that salient features of performances in those situations can be recognised and evaluated.*
>
> (Mislevy et al., 2012, pp. 35–37)

In parallel with reforms based on the promotion of particular assessment methods, there is considerable enthusiasm among some educators for the use of assessment 'rubrics'. However, once again, this enthusiasm often is limited to just this component of the assessment process, in isolation from other key assessment considerations. Rubrics are often designed only to grade isolated pieces of student work rather than to assemble evidence that can be used to draw inferences about progress within larger domains of learning.

At the present time, there are websites dedicated to the development and use of rubrics. These websites provide access to rubrics that teachers can download and use (for example, for use in assessing primary students' attempts at mathematics problems or persuasive writing tasks). Software has been developed to allow teachers to upload and share their own rubrics. These generally worded rubrics usually describe what teachers should look for in assessing student work. In the absence of well-defined learning domains, many rubrics attempt to fill this gap by describing the nature of improvement within a learning domain. However, these rubrics are often worded very generally. For example, the levels of a rubric may be described in almost identical terms, with distinctions between levels depending on the interpretation of graded adjectives such as 'very limited', 'limited', 'sound' and 'excellent'.

By contrast, in a Learning Assessment System, rubrics are informed by, and developed from, substantive descriptions of the nature of learning progress within the domain. The levels of a task rubric are instantiations of the domain itself and so provide the essential link that enables

them to contribute to inferences about where learners are in their learning within that domain. In other words, rubric development is intimately linked to an understanding of the domain. In much current assessment practice (for example, much test development) this link is implicit. A challenge for schools, education systems and practitioners is to build more explicit links between task rubrics and described learning domains in all assessment contexts.

The final step in any assessment process is to draw a conclusion from assembled assessment evidence. When assessments are undertaken during teaching, assessment conclusions are usually focused on how well students are mastering the taught content. At the end of a course, it is common to aggregate marks or grades to obtain an overall result. The general intention is to indicate how well students have performed overall in the course and to convey this as a percentage, course grade or percentile. Assessment conclusions of these kinds are not attempts to identify in an interpretable way where students are in their learning progress within a domain. Percentiles and 'grading on the curve' make no attempt at substantive interpretation of learning; percentages typically are uninterpretable because they are confounded by task difficulties; and most grading schemes are more focused on judging and comparing student performances than on communicating where individuals are in their learning progress. Such conclusions provide limited useful information about the learning that has occurred.

As discussed in Section 3, modern measurement theory does make an attempt to infer, from task performances, where students are on an underlying proficiency scale and so to provide substantive interpretations of students' assessment performances or responses. However, modern measurement theory has been primarily focused on ensuring valid and reliable inferences from task performances – in other words, it has been focused on only one part of a complete learning assessment system. Measurement theory assumes that tasks have been developed to address some well-understood learning domain, but the proficiency scales that result are 'post hoc' in the sense that they are constructed from assessment data and are usually not related to, or tested against, prior conceptualisations of the learning domain.

The general challenge here is to promote the use of more coherent assessment systems in educational practice. By bringing together the sometimes separated components of an assessment system and focusing them on the common purpose of establishing where learners are in their progress within an empirically based learning domain, assessment conclusions are likely to be more valid and more useful in decision-making.

Building assessment literacy

At its heart, this review paper has argued for the development of much stronger connections between scientifically based understandings of learning and the assessment of learning progress. It also has argued for the use of educational assessment to understand rather than to judge learning, for the conceptualisation of assessment as an integral part of effective teaching and learning (rather than as something that stands apart from, and follows, teaching and learning) and for designing educational assessments around a coherent learning assessment 'system', the elements of which work together to explore where learners are in their learning.

Developing and implementing such changes in assessment thinking and practice constitute a long-term agenda. They will occur only over a number of decades. The reasons for this, in part, are widespread and deeply entrenched public conceptions of assessment, which are strongly aligned with traditional modes of educational delivery. As long as students are grouped by age and taught a common year-level curriculum, educational assessments are likely to be predominantly judgements and comparisons of student success.

Other reasons include deeply entrenched assessment practices. Agencies currently exist to conduct assessments for the traditional purpose of judging and comparing student success. This assessment purpose, and even the reporting formats used in pursuing it, can be enshrined in government legislation and bureaucratic practice. And even beyond this, substantial reform is unlikely while the field of educational assessment itself is divided and in disarray, and while levels of assessment literacy are low across the education community. Significant assessment

reform will depend on addressing each of these underlying impediments, beginning immediately.

University education faculties have an obvious role to play in building future levels of assessment literacy. However, with some notable exceptions, initial teacher education courses currently provide quite low-level treatments of assessment, usually based on 20th century introductory textbook concepts and distinctions, many of which hamper rather than promote clear thinking about assessment. At their worst, teacher education courses promulgate simplistic and inaccurate views of assessment that are unhelpful to expert professional practice (for example, assertions that externally developed tests 'de-skill' teachers).

The development of higher levels of assessment literacy will require a multi-pronged approach. Examples of high-quality assessment resources and practices have an important role to play in modelling the systematic assessment of learning and in showing what is possible. Inherent in such examples must be solid, research-based understandings of specific learning domains. The purpose of these examples must be to establish and communicate where learners are in their learning and the progress they make over time. Grappling with the theoretical underpinnings of such examples is likely to be challenging but rewarding for educators. In addition, new and more informative ways of communicating the results of assessment processes will be required. These high-quality examples should include assessments of foundational and other life skills and attributes that develop over extended periods of time.

Assessment literacy also can be built through high-quality continuing professional development, post initial teacher education (Griffin, 2010). In-school learning, based on networking and the sharing of practice with colleagues, is an important form of professional learning. But there is also, and will continue to be, a need for high-quality, high-level courses that build expert knowledge. Significant assessment reform will depend on the development of high levels of assessment expertise among some senior educators who can then coach and mentor colleagues. Consideration should be given to recognising individuals who develop such expertise – for example, through accreditation as specialists in educational assessment. The recruitment of accredited specialists in educational assessment to teach in initial teacher education programs would then improve the quality of teacher preparation in this essential aspect of professional practice.

Concluding comments

Assessment is an integral and essential component of effective learning, teaching and educational decision-making. Improvements in the quality of assessment information have the potential to enhance the effectiveness of the decisions made by teachers, educational leaders, parents and learners themselves, resulting in better learning and better educational outcomes.

This review paper has argued that the key to better assessment information is the reconceptualisation of assessment as the process of establishing where learners are in their learning within a learning domain at the time of assessment. This process can be undertaken at various levels of diagnostic detail to identify starting points for action and to monitor learning progress over time. All elements of this process need to work together as a 'system' to establish and provide feedback on where learners are in their long-term learning progress.

The widespread adoption of this conceptualisation of assessment will not be achieved easily. There are deeply entrenched public conceptions of assessment as the process of judging and grading students on how well they have learnt what they have been taught. And, within the education community, rather than seeing all assessment as having the same fundamental purpose, it has become popular to refer to the 'multiple purposes' of assessment and to assume that these multiple purposes require quite different approaches and methods of assessment. Despite the challenges, the reform of educational assessment is urgent.

List of ACER Research Conference 2009 papers

At the conference, five keynote, nine concurrent papers, and six poster sessions were presented. Downloads of papers presented at the conference, or synopses, are available at http://research.acer.edu.au/research_conference/RC2009/.

Keynote papers

Forster, M. (2009). Informative assessment – Understanding and guiding learning.

Gardner, J. (2009). Assessment for teaching: The half-way house.

Masters, G. N. (2009b). Assessment for teaching.

Scheinin, P. (2009). Using student assessment to improve teaching and educational policy.

Wildy, H. (2009). Making local meaning from national assessment data.

Concurrent papers

Anderson, P. (2009). What makes a difference? How measuring the non-academic outcomes of schooling can help guide school practice.

Cook, J. (2009). An Even Start: Innovative resources to support teachers to better monitor and better support students measured below benchmark.

Humphry, S., & Heldsinger, S. (2009). Do rubrics help to inform and direct teaching practices?

Klenowski, V., & Gertz, T. (2009). Culture-fair assessment: Addressing equity issues in the context of primary mathematics teaching and learning.

Mendelovits, J., & Searle, D. (2009). PISA for teachers: Interpreting and using information from an international reading assessment in the classroom.

Spencer, K., & Balacco, D. (2009). Next practice: What we are learning about teaching from student data.

Timperley, H. (2009). Using assessment data for improving teaching practice.

Titmanis, P. (2009). Reflections on the validity of using results from large-scale assessments at the school level.

Wasson, D. (2009). Large cohort testing – How can we use assessment data to effect school and system improvement?

Posters

Conroy, D. (2009). A sporting chance for Aboriginal students in western NSW.

Corrie, T. (2009). On Track.

Cuttance, P. (2009). Prolearning: A real-time performance Information system for schools.

Harkness, A. (2009). Using internal school review data at school and system level to inform improvements in student learning – an online web-based application.

Latu, E. (2009). Effectiveness of feedback in mathematics learning.

McGuiness, R. (2009). How assessment effects children's learning.

References

Ainley, J., Fraillon, J., Gebhardt, E., & Schulz, W. (2012). *National Assessment Program – ICT Literacy Years 6 and 10 Report.* Sydney: Australian Curriculum, Assessment and Reporting Authority.

Ausubel, D. P. (1968). *Educational psychology, A cognitive view.* New York: Holt, Rinehart and Winston.

Australian Council for Educational Research. (2008). *Progressive achievement tests in reading*. (4th Edn.). Melbourne: ACER.

Australian Council for Educational Research. (2003). *TORCH Tests of Reading Comprehension*. Melbourne: ACER.

Australian Curriculum, Assessment and Reporting Authority. (2012). *The Australian Curriculum, English, Year 3* (Version 4.0). Sydney: ACARA.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills*. New York: Springer.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, *5*(1), 7–73.

Bloom, Benjamin S. (1968). *Learning for mastery*. Los Angeles, CA: University of California Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school: Expanded edition*. Washington: National Research Council.

Catley, K., Reiser, B., & Lehrer, R. (2005). *Tracing a prospective learning progression for developing understanding of evolution.* Commissioned paper prepared for the National Research Council's Committee on Test Design for K–12 Science Achievement, Washington, DC.

Chall, J. S. (1996). *Stages of reading development*. Orlando, FL: Harcourt Brace & Company.

Clauser, B. E., Margolis, M. J., Clyman, S. G., & Ross, L. P. (1997). Development of automated scoring algorithms for complex performance assessments. *Journal of Educational Measurement*, *34*, 141–161.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*(6), 1047–1085.

Dawson, T. L., & Stein, Z. (2008). Cycles of research and application in education: Learning pathways for energy concepts. *Mind, Brain, and Education*, 2, 89–102.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.

Education Department of Western Australia. (1994). *First steps spelling developmental continuum*. Melbourne: Longman Cheshire.

Education Department of Western Australia. (1997). *First steps reading developmental continuum*. Perth: Rigby Heinemann.

Fullan, M., Hill, P. W., & Crévola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.

Griffin, P. (2010), Developmental assessment: lifting literacy through professional learning teams. *Assessment in Education*, 17, 383–397.

Griffin, P. (2013). As reported by J. Topsfield in 'Results flatline for top students', *The Age*, 10 January 2013. Accessed January 11, 2013 from www.theage.com.au/national/education/results-flatline-for-top-students-20130109-2cgud.html

Griffin, P., McGaw, B., & Care, E. (2012). (Eds.). *Assessment and teaching of 21st century skills*. New York: Springer.

Harlen, W. (1997). *Making sense of the research on ability grouping*. Edinburgh: The Scottish Council for Research in Education.

Hattie, J. (2003). Teachers make a difference: What is the research evidence? Paper presented at ACER Research Conference *Building Teacher Quality: What Does the Research Tell Us?* 19–21 October 2003, Melbourne.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Hauser, C. (2003). So, what d'ya expect? Pursuing individual student growth targets to improve accountability systems. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Kanizsa, G. (1955): Margini quasi-percettivi in campi con stimolazione omogenea. *Rivista di Psicologia 49,* 7–30. (Translated into English by W. Gerbino 1987 as Quasi-perceptual margins in homogeneously stimulated fields. In S. Petry & G. E. Meyer (Eds.), *The perception of illusory contours*. New York: Springer. pp. 40–49.)

Kenney, J. F. (1997). New testing methodologies for the Architect Registration Examination. *CLEAR Exam Review*, *8*(2), 23–28.

Khoo, S. T. (2008). Statistical modelling of assessment and educational data from a progress monitoring perspective. Invited presentation to the Statistical Society of Australia, Victoria Branch.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington: National Research Council.

Marshak, D. (2003). No child left behind: A foolish race into the past. *Phi Delta Kappan*, Nov. 2003, 229–31.

Marton, F. (1986). Phenomenography: A research approach to investigating different understandings of reality. *Journal of Thought*, *21*(3), 28–49.

Masters, G. N. (2009). *Understanding and leading learning*. Presentation to principals' Big Day Out, Victorian Department of Education and Early Childhood Development, Melbourne, 1 August 2008.

Masters, G. N., & Forster, M. (1997a). *Mapping literacy achievement: Results of the National School English Literacy Survey*. Canberra: Department of Education, Training and Youth Affairs.

Masters, G. N., & Forster, M. (1997b). *Developmental assessment*. Assessment resource kit (ARK) for Teachers. Melbourne: ACER.

McDermott, L. C. (1993). Guest comment: How we teach and how students learn – a mismatch? *American Journal of Physics*, *61*, 295–298.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13–23.

Ministerial Council on Education, Employment, Training and Youth Affairs (2008). *Melbourne Declaration on Educational Goals for Young Australians*. Melbourne: MCEETYA. Accessed February 1, 2013 from http://www.mceecdya.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf

Mislevy, R., Behrens, J. T., Dicerbo, K. E., & Levy, R. Design and discovery in educational assessment: evidence-centred design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, *4*(1), 11–48.

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better.* London: McKinsey & Company.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: US Department of Education.

Nussbaum, J. (1979). Children's conceptions of the earth as a cosmic body: A cross-age study. *Science Education*, *63*(3), 83–93.

Organisation for Economic Co-operation and Development. (2007). *Understanding the brain: The birth of a learning science*. Paris: OECD.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Pellegrino, J. W., & Hilton, M. L. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington: National Academy of Sciences.

Quellmalz, E. S., Timms, M. J., & Schneider, S. A. (2009). *Assessment of Student Learning in Science Simulations and Games*. Unpublished paper, WestEd.

Ramsden, P., Masters, G., Stephanou, A., Walsh, E., Martin, E., Laurillard, D., & Marton, F. (1993) Phenomenographic research and the measurement of understanding: An investigation of students' conceptions of speed, distance, and time. *International Journal of Educational Research*, *19*(3): 301–16.

Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education*. New York: Springer Science+Business Media.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. New York, NY: Kluwer Academic Publishers.

Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). *ICCS 2009 International Report: Civic knowledge, attitudes, and engagement among lower secondary school students*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Thomson, S., Wernert, N., Underwood, C., & Nicholas, M. (2008). *TIMSS 07: Taking a closer look at mathematics and science in Australia*. Melbourne: ACER.

Thomson, S., De Bortoli, L., Nicholas, M., Hillman, K., & Buckley, S. (2011). *Challenges for Australian education: Results from PISA 2009*. Melbourne: ACER.

Thomson, S., & De Bortoli, L. (2012). *Preparing Australian students for the digital world.* Melbourne: ACER.

Timms, M. J. (2010). Using item response theory to select hints in an intelligent tutoring system. *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building technology rich learning contexts that work*, 213–221.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Cambridge, MA: Harvard University Press.

Wiliam, D. (2007). Once you know what they've learned, what do you do next? Designing curriculum and assessment for growth. In R. Lissitz (Ed.), *Assessing and modeling cognitive development in schools*. Maple Grove, MN: JAM Press.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M. (2005). *Constructing measures: An item-response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M. (2009, December). Assessment for learning and for accountability. Paper presented at the *Exploratory Seminar: Measurement Challenges within the Race to the Top Agenda*, Princeton, NJ.